THE
QUARTERLY JOURNAL
OF
EXPERIMENTAL PSYCHOLOGY

EDITORIAL BOARD:

R. J. AUDLEY
A. D. BADDELEY
G. S. BRINDLEY
D. E. BROADBENT
J. BROWN
V. R. CANE
R. DAVIS
R. L. GREGORY
I. M. L. HUNTER
H. KAY
N. MORAY
P. M. A. RABBITT
R. L. REID
N. S. SUTHERLAND
M. TREISMAN
A. M. UTTLEY
P. H. VENABLES
D. M. VOWLES
L. WEISKRANTZ
A. T. WELFORD

Editor: A. J. WATSON

Assistant Editor: KATHLEEN P. WATTS

VOLUME XX, 1968

PUBLISHED FOR
THE EXPERIMENTAL PSYCHOLOGY SOCIETY
BY
W. HEFFER & SONS LTD. CAMBRID

The Quarterly Journal of Experimental Psychology

Volume XX

CONTENTS

J. G. Wolff, J. Delacour, R. H. S. Carpenter and G. S. Brindley. The patterns seen when alternating electric current is passed through the eye 1

G. C. Grindley and V. Townsend. Voluntary attention in peripheral vision, and its effects on acuity and differential thresholds 11

G. R. Hardy and D. Legge. Cross-modal induction of changes in sensory thresholds 20

P. H. Venables and L. A. Warwick-Evans. The effect of stimulus amplitude on the threshold of fusion of paired flashes 30

L. H. Shaffer. Refractoriness in information processing 38

M. C. Corballis, W. Lieberman and D. Bindra. Discriminability and central intermittency in same-different judgements 51

J. M. von Wright. Selection in visual immediate memory 62

W. R. Garner and R. L. Gottwald. The perception and learning of temporal patterns 97

N. McK. Agnew, S. Pyke and V. Pezer. Frequency and recency effects on response biasing in pseudo-perceptual tasks 110

J. L. Bradshaw. Pupil size and problem solving 116

H. B. Savin. On Conrad's prefix and grouping in short-term memory 123

H. H. Clark and E. V. Clark. Semantic distinctions and memory for complex sentences 129

A. Treisman and G. Geffen. Selective attention and cerebral dominance in perceiving and responding to speech messages 139

S. A. Pizzurro and J. Van Laer. One-trial learning with control of item-difficulty 151

M. R. Harter and C. T. White. Periodicity within reaction time distributions and electromyograms 157

L. Karlin and R. Kestenbaum. Effects of number of alternatives on the psychological refractory period 167

P. M. A. Rabbitt Three kinds of error-signalling responses in a serial choice task 179

R. L. DORNBUSH. Shadowing in bisensory memory - - - - - - 225

P. M. A. RABBITT. Repetition effects and signal classification strategies in serial choice-response tasks - - - - - - - - - 232

P. M. A. RABBITT. Channel-capacity, intelligibility and immediate memory 241

A. D. BADDELEY. How does acoustic similarity influence short-term memory? - - - - - - - - - - - - - 249

P. WRIGHT. Sentence retention and transformation theory - - - - 265

P. C. WASON. Reasoning about a rule - - - - - - - - - 273

K. A. PROVINS and D. J. GLENCROSS. Handwriting, typewriting and handedness - - - - - - - - - - - - - - 282

C. T. SURRIDGE, K. R. MOCK and A. AMSEL. Effect of interpolated extinction after partial delay of reward training on subsequent reacquisition and extinction - - - - - - - - - - 321

M. MILLODOT. Influence of accommodation on the viewing of an optical illusion 329

D. W. J. CORCORAN, D. D. DORFMAN and D. L. WEENING. Perceptual independence in the perception of speech - - - - - - 336

D. PRITCHATT. An investigation into some of the underlying associative verbal processes of the Stroop Colour Effect - - - - - 351

L. H. SHAFFER and J. HARDWICK. Typing performance as a function of text - - - - - - - - - - - - 360

SHORTER ARTICLES AND NOTES

P. N. JOHNSON-LAIRD. The interpretation of the passive voice - - - - 69

S. JONES. Instructions, self-instructions and performance - - - - 74

B. B. MURDOCK. Response latencies in short-term memory - - - 79

D. W. J. CORCORAN and D. L. WEENING. Acoustic factors in visual search - 83

D. LESTER and A. W. STILL. Comment and Further Discussion on "Spontaneous alternation and pattern of reinforcement" - - - - 86

R. N. HUGHES. Effects of age on novelty reactions and exploration in rats - 189

D. LIKELY and S. B. SCHNITZER. Dependence of the overtraining extinction effect on attention to runway cues - - - - - - - 193

R. WONG and J. L. JAMIESON. Infantile handling and the facilitation of discrimination and reversal learning - - - - - - - 197

W. B. TEMPLETON and F. A. GREEN. Chance results in utrocular discrimination - - - - - - - - - - - - - 200

R. W. SEKULER and K. HOULIHAN. Discrimination of mirror-images: choice time analysis of human adult performance - - - - - 204

D. R. Haslam. Some anomalies in the study of experimental pain - - 208

G. H. Fisher. Gradients of distortion seen in the context of the Ponzo illusion and other contours - - - - - - - - - 212

R. H. Day and N. J. Wade. Involvement of neck proprioceptive system in visual after-effect from prolonged head tilt - - - - - - 290

H. J. Jeffrey. An inquiry into inappropriate constancy scaling - - 294

P. Bertelson and F. Tisseyre. The time-course of preparation with regular and irregular foreperiods - - - - - - - - - 297

G. Fischer, W. Viney, J. Knight and N. Johnson. Response decrement as a function of effort - - - - - - - - - - 301

H. A. Kuechler and P. C. Dodwell. Auditory signal detectability as a function of pre-experimental shock - - - - - - - 305

D. Kahneman, L. Onuska and R. E. Wolman. Effects of grouping on the pupillary response in a short-term memory task - - - - - 309

A. F. Sanders and J. J. F. Schroots. Cognitive categories and memory span:

I. Shifting between categories - - - - - - - - 370

II. The effect of temporal vs. categorical recall - - - - - 373

G. von Sturmer, T. Wong and M. Coltheart. Distraction and time estimation - - - - - - - - - - - - - 380

R. F. Garton and L. R. Allen. Familiarity and word recognition - - 385

R. A. Kennedy and A. L. Wilkes. Response-times at different positions within a sentence - - - - - - - - - - 390

W. A. Hershberger, P. R. Trantina and K. Cosgrove. Letter-naming time as a function of set familiarity and symbol distinctiveness - - 395

M. H. Sheldon. The effect of electric shock on rats' choice between familiar and unfamiliar maze arms: a replication - - - - - - 400

V. K. Kothurkar. Effect of similarity of interpolated materials on short-term recall - - - - - - - - - - - - 405

Book Reviews - - - - - - - - - 92, 218, 312 and 409

Publications Received - - - - - - - - - 94, 224, 320 and 414

Proceedings of the Experimental Psychology Society, 1967 - - - - 95

Index to Authors - - - - - - - - - - - - 415

THE QUARTERLY JOURNAL OF
EXPERIMENTAL PSYCHOLOGY

Vol. XX 1968 Part 1

THE PATTERNS SEEN WHEN ALTERNATING ELECTRIC CURRENT IS PASSED THROUGH THE EYE

BY

J. G. WOLFF, J. DELACOUR, R. H. S. CARPENTER and G. S. BRINDLEY

From the Physiological Laboratory, University of Cambridge

When alternating current of frequency up to 210 c/s. is passed through the eye while the retina is brightly and uniformly illuminated, conspicuous spatial patterns are seen. Other such patterns are seen when current of frequency up to 100 c/s. is passed through the eye in darkness. These patterns are here described, and their origin is discussed.

INTRODUCTION

Lohmann (1940), in a paper mainly concerned with other aspects of the sensitivity of the eye to electric currents, gave a brief description of patterns that could be seen when all or most of one retina was uniformly illuminated, and at the same time alternating electric current was passed through it. We can find no other published reference to these very striking and beautiful patterns except for passing mention in two papers (again mainly concerned with other topics) by one of us (Brindley, 1962, 1964).

Over a narrower range of frequencies, conspicuous patterns can be produced by passing alternating current through the eye in darkness. These were first described by Brindley (1955, p. 196) but only for 50 c/s., a frequency that is relatively unfavourable for producing them.

In the present paper we shall describe these phenomena more thoroughly, and speculate a little on their origin.

METHOD

Light and electricity were put into the eye through an eye-bath filled with 0·9 per cent. sodium chloride solution, as in the experiments of Brindley (1962, 1964). The indifferent electrode was a large wet aluminium plate, on which one hand was placed. The central 40° of the visual field was uniformly illuminated, usually at 220 cd./m.^{2} with white light. Where any other luminance or colour was used, this is explicitly stated in the text. The far periphery of the visual field was illuminated less brightly, by a factor of roughly 3. The electric currents were sinusoidal except where otherwise stated. Except where otherwise stated, 2 V. r.m.s. was applied between the eye and the hand. The current flowing through the subject was roughly 0·7 mA./V. at all the frequencies used.

The conjunctival sac of the eye used was anaesthetized with either 1 per cent. amethocaine or 2 per cent. cocaine.

The other eye was usually kept open and illuminated, but covering it made little or no difference to the patterns seen.

OBSERVATIONS

A. Retina uniformly illuminated

The following account is arranged in increasing order of frequencies of alternation of the electrical stimulus, and is divided into four stages. The beginnings of the second, third and fourth stages are defined by the appearance at or near the point of regard of a new patch of mauve or blue that tends to expand and fill the whole visual field as the frequency of alternation of the current is increased. A rather similar patch appears during the first stage, but it is fainter, less well defined, and different in colour.

The letters *w*, *d*, *c* and *b* stand for the four authors.

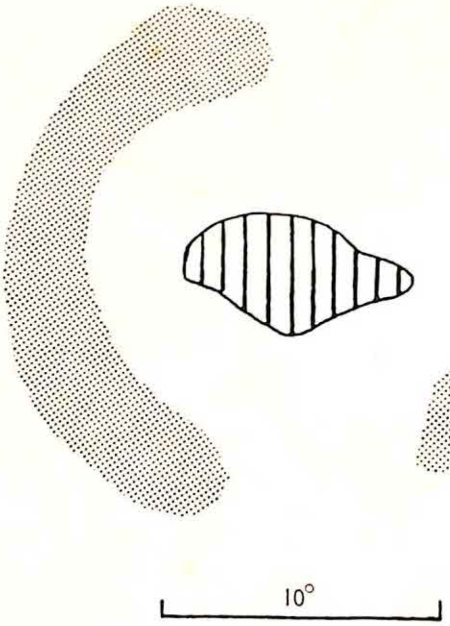
Stage 1

(i) At frequencies between 10 and about 40 c/s. the whole field flickers strongly, and in this flickering field vague and elusive patterns are seen that vary greatly from subject to subject, but are repeatable for any one subject. *b:* a regular horizontal and vertical chessboard. *c:* like interference patterns on a ripple tank; superimposed on these, a large-scale pattern of slowly-moving grey continents. *w:* a complex pattern of ill-defined light grey stripes.

When we showed these phenomena to Dr. J. R. Smythies, he commented that at frequencies between 10 and 40 c/s. they were just like those with which he was very familiar in flickering light (Smythies, 1959*a*, *b*); but the patterns seen at frequencies over 40 c/s. were something quite new in his experience.

(ii) At 30–35 c/s. a roughly elliptical patch of dirty yellow-green (*c*) or neutral grey (*b*, *d*) appears, centred on the point of regard. With increasing frequency it expands, until at 40 c/s. it is about 7° wide and 5° high. *d:* Figure 1; the patch consists at 40 c/s. of two shadow-like arcs. At higher frequencies it is difficult or impossible to see.

FIGURE 1



The pattern seen at 40 c/s. by *d*. In the centre (striped) is the blue or mauve patch (i) of Stage 2; surrounding it (speckled) are the neutral grey arcs (ii) of Stage 1.

(iii) From 10 to 50 c/s. the region of the blind spot looks different from the rest of the field. *c:* more intensely flickering, but not on the average brighter or darker. *b:* the special appearance of the blind spot is restricted to its boundary (as shown in Fig. 5 of Brindley, 1955); this boundary flickers more intensely than the rest of the field, and is brighter.

Stage 2

(i) At 40 c/s. (*b*, *c*, *d;* for *w* 48 c/s.) a blue or mauve patch appears, centred on the point of regard, and expands with increasing frequency (Fig. 1). This is the "spot" described on p. 517 of Brindley (1964).

(ii) By 50 c/s. (*b*, *c*, *d;* for *w* 62 c/s.) the patch is an irregular oval about 10° wide by 7° high.

(iii) It is bounded externally by a fine line of luminous cream (*c*) or turquoise (*d*). *b* and *w* see no distinct bounding line, but merely a sudden change from white or yellowish white outside to mauve or blue inside.

(iv) The centre of the patch is paler than its outer part (*b, c, d, w*).

(v) The patch flickers less strongly than the background near it (*b, c, d, w*).

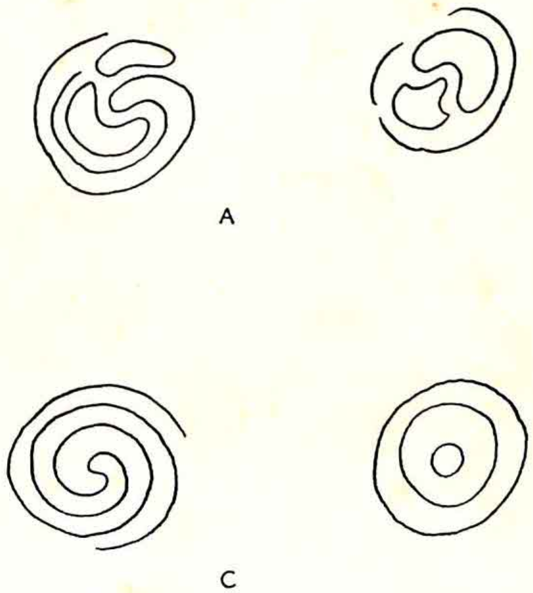
(vi) As the patch expands with increasing frequency, a new region of strong flicker appears at the centre of it (*b, c, d, w*).

(vii) At about 50 c/s. this central flickering area may be seen as a stationary pattern (*b, c, d*). *b:* from 50 to 55 c/s. straight radial stripes; from 55 to 60 c/s. these become curved like irregular daisy petals, but remain stationary; above 60 c/s. the radial "petals" move and become part of the pattern of yellow lines (see x below). *c:* at 60 c/s. a watered silk pattern, with a bright turquoise spot at the point of regard.

(viii) As the patch expands, areas of a similar colour appear in the far periphery and expand inwards and merge with it, first in the lower nasal quadrant (*b, c, w;* for *d* upper nasal), and then elsewhere.

(ix) The central and peripheral mauve or blue areas continue to merge until at 60 c/s. (*b, d;* for *c* 75 and for *w* 90 c/s.) all that remains of the former background is a crescent in the upper temporal (*b, c, w;* for *d* lower temporal) part of the field.

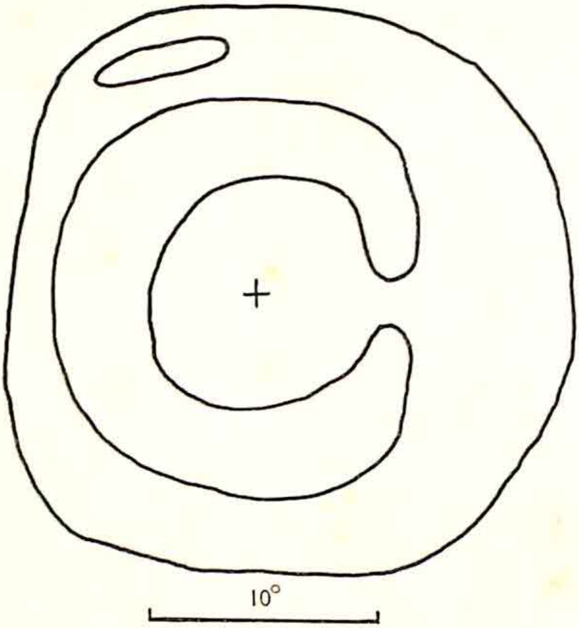
FIGURE 2



Four sketches of (x) of Stage 2 at 60 c/s. as seen by *d*. The lines are yellow (occasionally yellowish-green or orange), and about 3′ in thickness.

(x) At about the frequency at which the central and peripheral mauve or blue areas begin to merge, curved continuously moving yellow lines appear close to the point of regard (*b, c, d, w*). These lines form various patterns within the central 10° of the field. (Figs. 2 and 3). When each line first appears, it either ends indefinitely in the peripheral field, or, more commonly, forms a closed loop several degrees in diameter. Every loop disappears by shrinking to nothing, but its shrinking may be interrupted by brief periods of expansion or of change of shape without change of size. For *w*, each loop lasts about 2 sec., and there are about three in the field at any one time. For *b* and *d*, a loop lasts on the average about 5 sec. For *c, d* and *w*, but not for *b*, a yellow line sometimes forms a spiral centred on the point of regard (Fig. 2C).

FIGURE 3



(x) of Stage 2 at 70 c/s. as seen by *w*. Here and in Figures 4 and 8 the small cross marks the point of regard, and is not part of the pattern.

(xi) The yellow lines are about 3′ in width. For *w* and *b* they have narrow dark borders. For *c* their boundaries are hazy at frequencies below about 80 c/s. but sharp at higher frequencies.

(xii) The area occupied by the yellow-line figures expands with increasing frequency.

Stage 3

(i) At 65 c/s. (*b, d;* for *c* 80 and for *w* 90 c/s.) a second blue patch, darker and more saturated than that of Stage 2, begins to show itself in the centre of the field. It expands with increasing frequency.

(ii) Where yellow lines cross the patch, they appear more conspicuous and less mobile than elsewhere (*b, c, d, w*). For *b* and *d* they also appear finer where they cross the patch (see Fig. 4).

(iii) For *b* and *d* the patch is 10° wide and 6° to 9° high at 75 c/s. For *c* it is 10° in diameter (width and height about equal) at 85 c/s. For *w* it is 10° wide and 8° high at 100 c/s.

(iv) As the patch expands, areas of similar colour appear in the far periphery and expand to meet it (*b, c, d, w*).

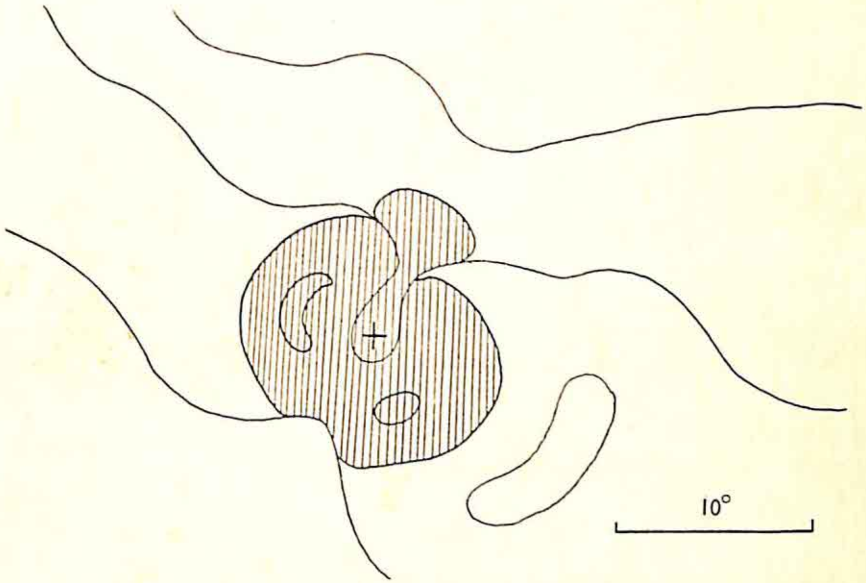
(v) The yellow-line patterns of (x) of Stage 2 remain conspicuous throughout Stage 3. The pattern seen by *d* at 85 c/s. is shown in Figure 5.

(vi) There is a faint nearly stationary flickering of light blue and dark blue in the central 10° of the field (*b, d, w*). This increases in size with increasing frequency; for *b* it has attained a diameter of 20° by 95 c/s., for *w* not until 130 c/s.

(vii) As the frequency is increased, the yellow lines cease to be concentrated in the central part of the field, long open-ended lines and large loops become commoner and small loops less common, and the pattern becomes roughly uniform throughout the whole field. For *d* the pattern attains this rough uniformity at 90 c/s., for *b* and *c* at 100 c/s., and for *w* at 130 c/s. Figure 2 of Brindley (1962) shows a typical pattern.

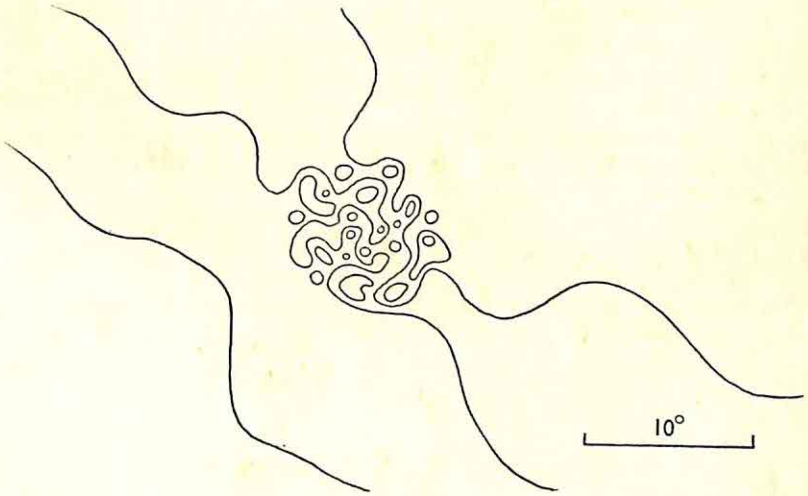
(viii) At 90 c/s., *d* sees a small turquoise spot with an irregular boundary flashing at the point of regard. *c* and *b* see a closely similar turquoise spot at very different frequencies, *c* at 60 c/s. (see (vii) of Stage 2) and *b* at 120 c/s. (see (ii) of Stage 4). *w* sees nothing of this kind.

FIGURE 4



Early Stage 3 (75 c/s.) as seen by *d*. The blue patch characteristic of Stage 3 is shown vertically striped.

FIGURE 5



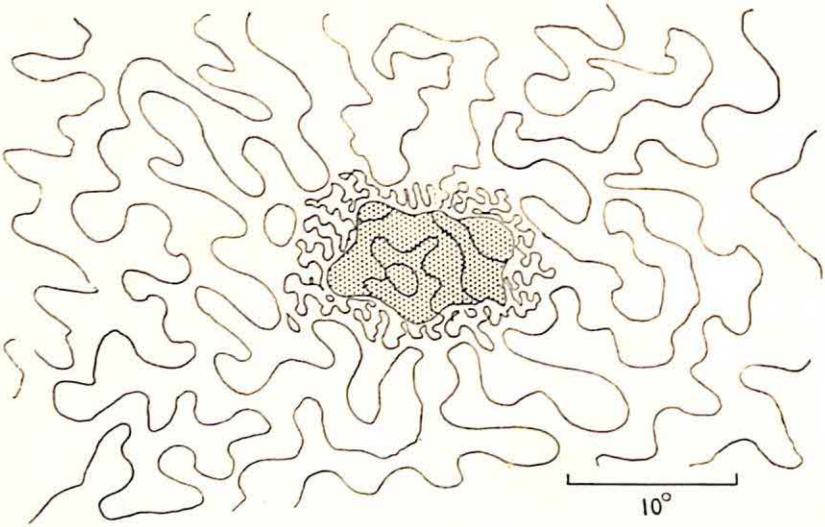
The yellow-line patterns seen at 85 c/s. by *d*.

Stage 4

w sees many of the phenomena of this stage only when the applied r.m.s. potential is increased to 2·5 V., and his descriptions below are for this potential. *b*, *c* and *d* continued to use 2 V.

(i) At 100 c/s. (*d*) a new blue patch appears in the centre of the field, still darker and more saturated than that of Stage 3. For *b* it appears at 110 c/s. and for *w* (2·5 V.) at 115 c/s. For *c* it is formed by expansion of the turquoise spot (vii) of Stage 2, but this expansion does not begin until 90 c/s.

FIGURE 6

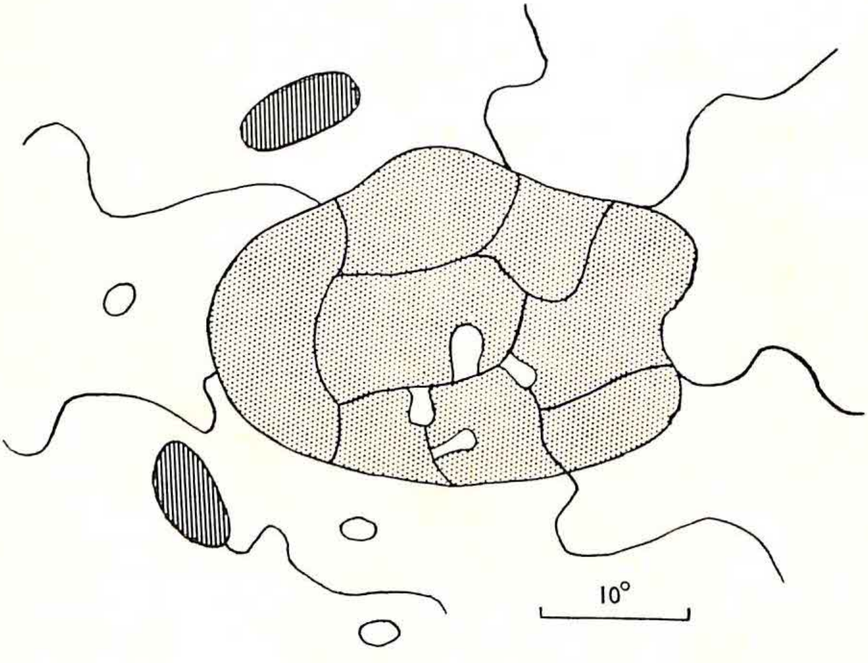


Early Stage 4 (110 c/s.) as seen by *d*.

(ii) The blue patch that defines Stage 4 shares, but in lesser degree, the tendency of the patches that define Stages 1, 2 and 3 to expand with increasing frequency. It differs from them in that it may appear as several separate pieces (*b*, *c*, *d*, *w*). For *b* it always does so; there are three, four or five pieces, one very small, centred on the point of regard, and greenish blue, the rest centred about 3° from the point of regard and violet-blue; a slight increase of frequency causes them to merge.

(iii) Patterns of yellow lines are still present throughout the field. Where yellow lines cross the central blue patch they become conspicuously thinner (*b*, *c*, *d*, *w*).

FIGURE 7



Middle Stage 4 (130 c/s.) as seen by *w*. The speckling represents uniform blue, but the two patches shown striped are actually seen striped.

(iv) Blobs of yellow appear within the central blue patch, and (for *c*, *d* and *w*) fuse with some of the yellow lines. *b* likens the yellow blobs to holes in the blue patch, through which the general background shows; for him they never fuse with yellow lines, and indeed are sometimes crossed by them, the lines appearing saturated yellow upon the less saturated yellow blobs. For *w* the blobs sometimes flash on and off.

(v) The patch remains roughly circular, expanding slightly with increasing frequency, but is continually beaten about by the lines and blobs, and occasionally forced into two or three separate parts (*b*, *c*, *d*, *w*).

(vi) At about 130 c/s. (*b*, *d*, *w;* 110 c/s. for *c*) other patches appear in the peripheral field. These consist of fine parallel blue and white stripes (*b*, *c*, *d*, *w;* see Fig. 7). The stripes are so fine that it is difficult to be sure of their direction, but *b* and *c* think that they are vertical.

(vii) For *b*, *c* and *d*, as the frequency is increased the central and peripheral patches merge. The peripheral patches seem to remain striped and the central one to remain unstriped, but the distinction becomes difficult to make, and no sharp boundary appears between striped and unstriped. For *w*, as the frequency is increased, the striped patches increase in number and size and the central patch breaks up; patches of the two kinds do not merge. For all of us yellow lines and yellow blobs are abundant at this stage, and move about rapidly. There are roughly equal amounts of yellow and blue in the field, and the general impression is of yellow sunlight playing on a gently rippling ultramarine lake.

(viii) Around 150 c/s. spots of a more saturated and darker blue come and go in the central 10° of the field, perhaps eight at any one time (*b*, *c*, *d*, *w*). *b* and *d* suggest that these might be considered to mark the beginning of an abortive fifth stage; *c* and *w* disagree.

(ix) For *d* and *c* between 140 and 160 c/s. the field loses its brilliance and the colours become darker. *b* and *w* do not see this.

(x) About 160 c/s. yellow begins to dominate and the blue to be more and more broken up (*b*, *c*, *d*, *w*). Figure 2B of Brindley (1962) shows what *b* sees at about this frequency and a rather higher luminance.

(xi) At 190 c/s. (*c*, *d*, *w;* 215 c/s. for *b*) the last of the blue patches disappears. For *d* and *b* the last surviving blue patch is always in the lower temporal part of the field, and for *w* near the point of regard.

(xii) From about 180 c/s., loops (often concentric) or horseshoes of pale grey line, reminiscent of smoke-rings, centred on the point of regard, expand slowly (*b*, *c*, *d*, *w;* see Fig. 8). Their slow expansion is practically independent of frequency over the range of frequencies at which they can be seen. Between 180 and 190 c/s. yellow lines and expanding loops of grey line may be present at the same time, and seem to be distinct and independent; above 190 c/s. there are grey lines but no yellow.

(xiii) The expanding loops of grey line fade as the frequency is increased, and above 210 c/s. (*d*, *c*; 220 for *b* and 225 for *w*) there is nothing to be seen.

B. *Waveform and strength of current, and colour and luminance of light*

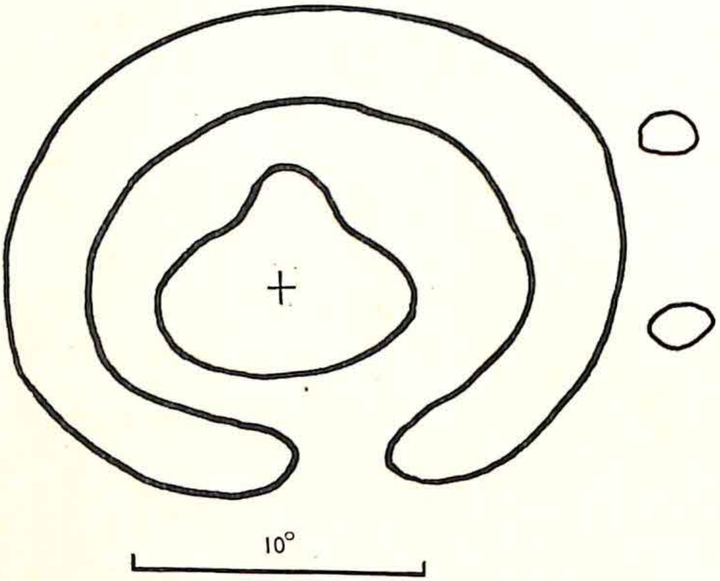
Square-wave current behaves nearly like sinusoidal current of the same frequency and slightly larger peak-to-peak amplitude.

Short pulses of current, alternating with much longer and weaker pulses in the opposite direction so as to prevent electrolysis of the tissues, can produce patterns with both directions of flow; but the two directions are not exactly equivalent.

If while any of the patterns of Section A is being observed the current is switched off, the pattern disappears instantly. The results of switching the current on after an interruption vary with the stage (i.e. frequency). In Stage 2, the pattern is restored instantly (*b*, *c*). In Stage 3, the central blue patch is restored instantly, but the number of yellow lines is for the first few (or many) seconds slightly reduced (*b*), or very much reduced (*c*). *c* finds that the lines are often quite absent when the current is switched on after an interruption and can remain absent for over a minute, returning only if he *blinks*. Momentary interruption of the light is not sufficient to bring back the yellow lines; closure of the lids is necessary. In Stage 4, the whole pattern is restored instantly on switching on the current after an interruption, but for the first few seconds after such restoration the area of blue in it is strikingly greater (and of yellow correspondingly less) than subsequently.

The patterns that we describe are seen best with the strongest currents that pain allows the subject to tolerate. The 2 V. r.m.s. (approximately 1·4 mA. r.m.s.) that we used for the observations of Section A is not far short of this but *w* and *b* were able to make some observations at 4 V. without seeing anything strikingly new. The range of frequencies at which each kind of pattern can be seen is increased slightly, both upward and downward, by increasing the strength of the stimulus, and the vividness of the patterns is generally enhanced.

FIGURE 8



Late Stage 4 (180 c/s.) as seen by *w*. The lines are pale grey, and a little broader and less well defined than the yellow ones seen at lower frequencies. The whole pattern expands slowly, new lines appearing every few seconds near the point of regard.

The luminance of the light can be increased or decreased tenfold from the 220 cd./m.^{2} that we habitually used without any substantial change in the patterns. We did not try lights brighter than 2,200 cd./m.^{2}. With very dim lights, patterns intermediate between those of Sections A and D can be obtained.

The colour of the light does not at all influence the shapes of the patterns seen. It influences their colours in a rather uninteresting way, the true and apparent colours being roughly additive.

C. *Tangential versus radial current*

One of us (*b*) examined the patterns produced by alternating current passed between electrodes in the upper and lower fornices of the conjunctival sac, the central 40° of the visual field being illuminated at 220 cd./m.^{2} with white light. Currents between 3 and 4 mA. r.m.s. were required to produce strong patterns, and these currents were painful even with thorough cocainization of the conjunctival sac. The patterns seen were similar in their main features to those of Section A of this paper, except that they occurred only in the upper and lower parts of the field, and were entirely absent from a horizontal band across the centre of the field, a band that could be caused to tilt by rotating the electrode-array about the visual axis. This is compatible with the supposition that the patterns of this paper are produced only by current that flows radially through the retina.

D. *Retina in darkness*

The patterns can be divided into three stages roughly corresponding to the first three stages of the patterns seen in the light.

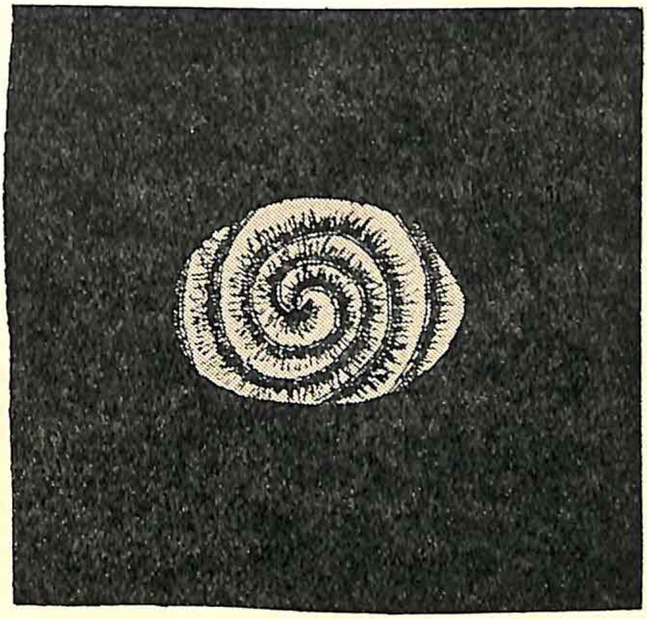
Stage 1

Between 10 and about 40 c/s. the whole field flickers strongly. At 45 c/s. *b* and *d* see a faint pattern of mainly horizontal wavy lines, alternately bright and dark, in the central 6° or so of the fields; *c* sees this, and also, in the same region of the field, small orange and blue patches moving about.

Stage 2

At about 50 c/s. a small bright lilac patch appears centred on the point of regard, and expands with increasing frequency. Between 55 and 70 c/s. this patch becomes blue rather than lilac, and takes on the very vivid regular spiral appearance illustrated in Figure 9 (*b*, *c*, *d*). The outwardly-directed dark teeth of this spiral flicker strongly (*b*, *c*, *d*). For all of us the spiral is sometimes right-handed, sometimes left-handed; it is rarely seen to reverse, and can remain stable for many minutes. Above 70 c/s. the spiral pattern becomes fainter and much less regular.

FIGURE 9



Pattern seen in darkness at 60 c/s. Drawing by *c*.

Stage 3

At about 70 c/s. a second small bright lilac patch appears centred on the point of regard, and, like that of Stage 2, expands with increasing frequency. The peripheral field remains faintly luminous, and both it and the central patch are crossed by dark lines that form moving patterns rather like those formed by yellow lines when the retina is illuminated (x, xi and xii of Stage 2 and ii, v and vii of Stage 3 of the light patterns).

No patterns are seen above 100 c/s. in darkness.

DISCUSSION

There is nothing that we can say confidently about the origin of the conspicuous and beautiful patterns that we have described, but we should like to put forward a few speculations.

The failure to produce these patterns with currents flowing through the retina parallel to its surface makes it likely that, as with ordinary electrical phosphenes (Brindley, 1955), the electricity is acting on structures oriented radially in the retina, i.e. bipolar cells or rods and cones. But this cannot be concluded as firmly for the patterns of this paper as for ordinary phosphenes. In experiments with two conjunctival electrodes we know that a density of tangential current can be produced which, if it were radial, would produce a phosphene in the dark, for in the fovea, where the "radial" elements depart appropriately from the strictly radial orientation, it does so (Fig. 4 of Brindley, 1955); but there is no similar check that this tangential current density is such that, if it were radial, it would produce the spatial patterns with which the present paper is concerned.

The near-constancy from subject to subject of very many details of these patterns from Stage 1 onward contrasts strikingly with the great variability of the patterns seen when (in the absence of electric current) the whole visual field is filled with light flickering at less than 40 c/s. (see Smythies, 1959). The constancy, and the irrelevance of stimuli to the other eye, provoke us to suggest that in some sense our patterns arise peripherally, in the retina rather than the brain. But such a suggestion needs to be elaborated a little before one can be quite clear what it means.

People who have seen the patterns that we are describing generally say that what they see is exactly what they would expect to see if a certain pattern of moving lines and patches were really displayed in front of them; they are of the opinion that it would be possible (though technically difficult) to make a cinematograph film that mimicked any of our patterns so exactly that one could not tell the difference.

From this probable identity of experience it seems reasonable to suppose that, at some late stage in the visual pathway, identical patterns of activity are produced by the electrical or electrical-and-optical stimulus and by the purely optical stimulus that (probably) mimics it. But it is exceedingly unlikely that there is any sort of mobile spatial patterning in the activity of the structures directly affected by the alternating current in our experiments, for the distribution of the current itself must be practically static. Thus in speculating that the Lohmann patterns are produced in the retina, we mean that already by the stage of optic nerve fibres the spatially almost unpatterned activity generated directly by the uniform light and electricity has been converted into a spatially patterned activity resembling that produced by a purely optical mimicking stimulus.

We are well aware that our speculation is uncertain and difficult to test.

References

BRINDLEY, G. S. (1955). The site of electrical stimulation of the human eye. *J. Physiol.*, **127**, 189–200.

BRINDLEY, G. S. (1962). Beats produced by simultaneous stimulation of the human eye with intermittent light and intermittent or alternating electric current. *J. Physiol.*, **164**, 157–67.

BRINDLEY, G. S. (1964). A new interaction of light and electricity in stimulating the human retina. *J. Physiol.*, **171**, 514–20.

LOHMANN, H. (1940). Über die Sichtbarkeitsgrenze und die optische Unterscheidbarkeit sinusförmiger Wechelströme. *Z. Sinnesphysiologie*, **69**, 27–40.

SMYTHIES, J. R. (1959*a*). The stroboscopic patterns. I. The dark phase. *Brit. J. Psychol.*, **50**, 106–16.

SMYTHIES, J. R. (1959*b*). The stroboscopic patterns. II. The phenomenology of the bright phase and after-images. *Brit. J. Psychol.*, **50**, 305–24.

Manuscript received 12th June, 1967.

VOLUNTARY ATTENTION IN PERIPHERAL VISION AND ITS EFFECTS ON ACUITY AND DIFFERENTIAL THRESHOLDS

BY

G. C. GRINDLEY and VALERIE TOWNSEND

From the Psychological Laboratory, Cambridge

The three experiments described in this paper were intended to show whether voluntary attention to a particular part of the peripheral visual field had any effect on the accuracy of the subject's perception. Test objects near to threshold value (for acuity or for changes in luminance) were used. The experiments were also designed to study the possible distracting effect of other stimuli presented simultaneously in different parts of the field.

In Experiment 1 the subject could be given foreknowledge of the position in which an acuity test object would appear. In Experiment 2 the signal of where to attend was given simultaneously with the exposure of an acuity test object. In Experiment 3 differential thresholds for luminance were investigated by methods similar to those used in Experiment 2.

It was found in all these cases that the instructions to attend to a particular part of the peripheral field had no significant effect on perception unless there were simultaneously exposed "competing" stimuli in other parts of the field. The results support the view that, in peripheral vision, attention acts selectively on the immediate memory trace only when there is a complex pattern of stimulation.

INTRODUCTION

An experiment by Helmholtz published in 1866 was one of the first to suggest that voluntary attention to one part of the visual field can, in some circumstances, make perception clearer there. He was looking at complicated pictures and he remarks. . . .

> "It is a curious fact that the observer may be gazing steadily at the fixation mark, and yet at the same time he can concentrate his attention on any part of the dark field he likes, so that when the spark comes he will get an impression about objects in that particular region only. In this experiment the attention is entirely independent of the position and accommodation of the eyes or, indeed, of any known variations in or on the organ of vision. Thus it is possible, simply by a conscious and voluntary effort, to focus the attention on some definite spot in an absolutely dark and featureless field. In the development of a theory of the attention, this is one of the most striking experiments that can be made." (Helmholtz, 1925.)

In this experiment by Helmholtz there was no control of eye movements though it now seems unlikely that they were important. The experiments can, however, be considered as the starting point of more recent work on the effect of attention in peripheral vision. The effect is so striking that it can easily be shown with any form of tachistoscope, or by short exposures of a complex group of figures on a screen to a group of students.

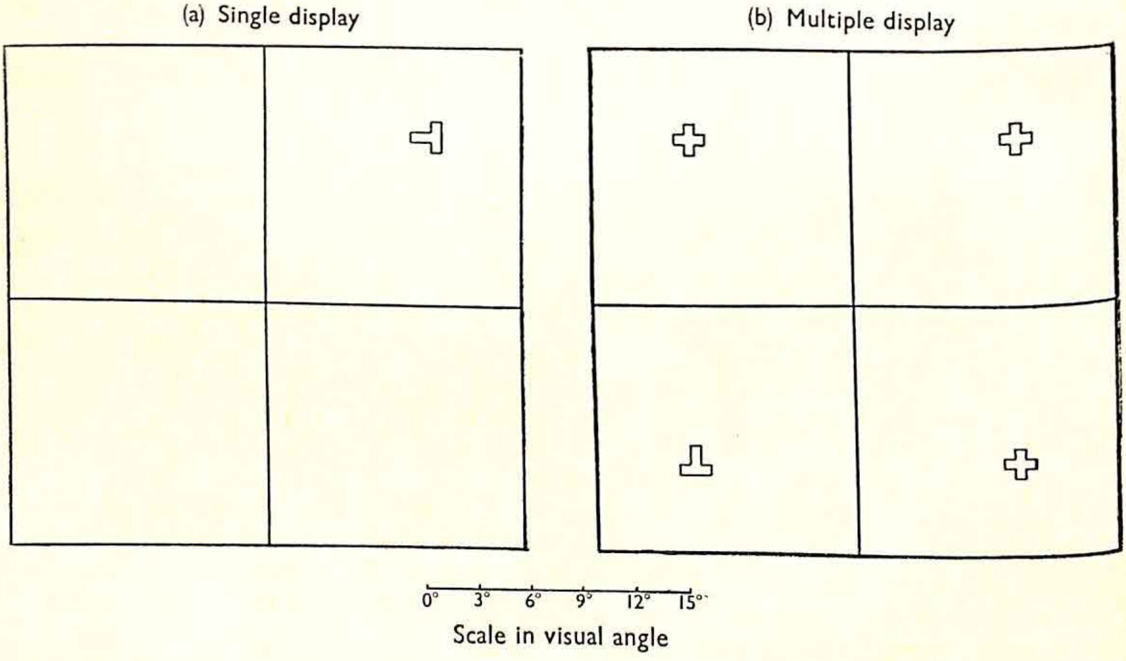
More recently Sperling (1960) and Averbach and Sperling (1960) have done some well controlled experiments in which a group of letters (greater than the subject's span of apprehension) was exposed tachistoscopically. They found that if there was no signal to attend to a particular part of the field and the subject was asked only to report as much as he could of what he had seen, the report would, on the average, usually give about four or five letters correctly. But if there was an "attention signal" just before or simultaneously with (or even just *after* the exposure) directing the subject to one part of the field, the subject could usually report the letters in that position quite correctly. The subject often added that he had "seen" letters in other places but had forgotten what they were. With attention directed to a different part of the same pattern of letters, similar results were found. This technique enables the effect of eye movements to be ignored. Averbach and Sperling conclude that under their conditions, attention acts on our immediate memory rather than on what we "see."

In the experiments described in this paper we were concerned with the effect of attention on briefly exposed peripheral test objects when these were near to threshold value. We have attempted to examine this question in the case of acuity and differential thresholds of brightness and have compared the results of foreknowledge of the position in which the test object is to appear with the case in which the attention signal is given simultaneously with the test object. We have also compared the case where only one test object is exposed with that in which there are also "competing stimuli" exposed in other parts of the peripheral field. (In each of the experiments there were, of course, control trials in which no attention signal was given.)

EXPERIMENT 1. *Acuity in the Peripheral Field, with and without foreknowledge of the place where the test object will appear.*

Apparatus. The subject sat with his head against a rest and viewed with his left eye a dimly illuminated translucent screen (0·25 ft. L.) at a distance of 20 in. from the eye. The field of the right eye was blacked out. A crosswire on the screen formed a central fixation point. Behind the screen there was a slide projector with a Compur shutter which threw figures (of luminance about 1·3 ft. L.) onto the screen for any desired exposure time. Two lights, one about 30° visual angle above the fixation point and another the same distance below, illuminated the subject's left eye sufficiently for its movements to be observed through a laboratory telescope. Saccadic movements of more than about 2° could be detected by this method; this was described in two of our earlier papers (Grindley and Townsend, 1965 and 1966).

FIGURE 1.



Typical displays used in Experiment 1.

Display. Two kinds of display were used which we shall call "single" and "multiple." An example of a single display is illustrated in Figure 1*a*, and of a multiple display in Figure 1*b*. In each case the critical test object was in the form of a T; which could be shown in any of four orientations and in any of the four quadrants into which the field was divided by the cross-wire. After each trial the subject was asked to state the quadrant in which the test object appeared and also its orientation. The verbal code used was a simple one about which none of the subjects had any difficulty. Thus the correct answer in Figure 1*a* would have been "2 left" and in Figure 1*b* "3 up." In all cases, as illustrated in Figures 1*a* and *b*, the objects presented were at about 13° visual angle from the fixation point. In the case of multiple displays the other objects were always crosses as illustrated in Figure 1*b* and they were all of the same area as the test object, i.e. each object would produce the same total change in luminous flux on its own region of the retina.

Procedure. In the actual experiment the subject was shown a series of 16 slides with a single test object followed by a series of 16 slides with competing objects or *vice versa.* He was told that in some trials he would be informed of the quadrant in which the T would appear (and this information was always correct), but he was never told which way up the test object would be and he was never given knowledge of his results. In fact in both series the number of times that the test object appeared in each quadrant was four, and in half the cases the subject was told the quadrant. He was not told beforehand either the number of exposures in each series or that the number of times that the test object appeared in each quadrant would be equal. None of the subjects appeared to have guessed this. Some preliminary trials were given to each subject both with the test object exposed alone and also with competing objects. This served two purposes, firstly, to allow the subject to know the "code" to be employed in his answers, and secondly, to give the experimenters some idea of the exposure time which would enable the subject to give answers which were better than chance under each condition, but not 100 per cent. correct. In these preliminary trials the subject was not told in which quadrant the test object would appear. In most cases it was found that the subject needed a longer exposure when there were competing objects than when the test object appeared alone.

In each trial the subject was asked to fixate the cross and if the experimenter merely said "fix" the subject replied "yes" when he was satisfied he was fixated. But if the trial was one with foreknowledge, the experimenter would say for example "fix 2," and the subject would say "yes" as soon as he was fixated on the cross and when he was attending to quadrant 2. One experimenter then immediately released the shutter while the other watched the subject's eye. As far as we could estimate this procedure under either condition resulted in a delay of about 1 to 2 sec. between "fix" and the exposure.

Results. It was found in the *Single* display that the 10 subjects gave a total of 36 correct judgements out of a possible 80 when they had foreknowledge of the quadrant in which the test object was to appear, while with no foreknowledge they gave a slightly greater number of correct answers—44. This difference on a Wilcoxon test was not significant at any level. In fact in a previous experiment using the same apparatus but with less well controlled conditions there was a slight but quite insignificant difference in the opposite direction.

Using *Multiple* displays, the results were quite different, namely that with foreknowledge, the subjects gave 66 correct responses out of a possible 80, while with no foreknowledge they gave only a total of 35. No statistical test was necessary since all 10 subjects gave a greater number of correct answers with foreknowledge than without.

EXPERIMENT 2. *Acuity with a Simultaneous Attention Signal.*

In Experiment 1 there was a remote possibility that the results might have been affected by changes in the adjustment of the eye which we could not detect (see Discussion). We therefore did an experiment using an attention signal given simultaneously with the brief exposure of the test object or objects. This method was similar in principle to that employed by Averbach and Sperling.

FIGURE 2

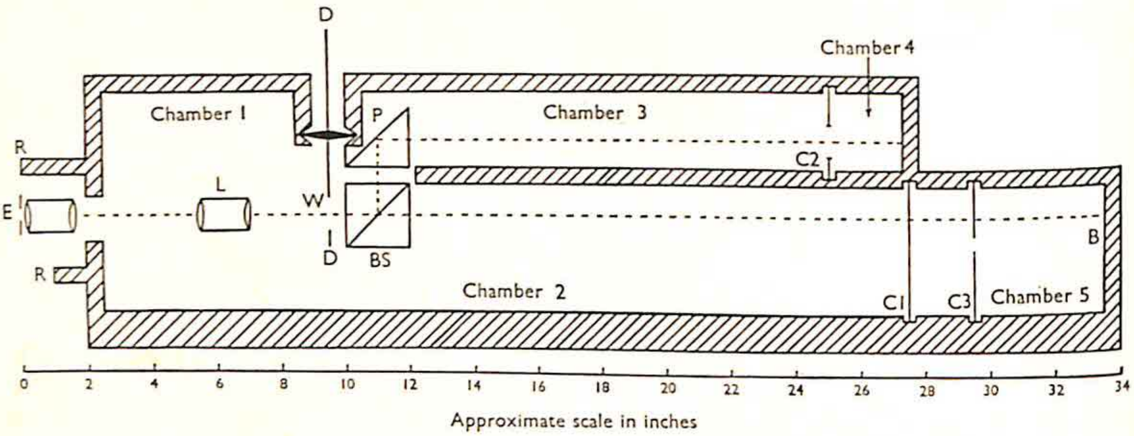


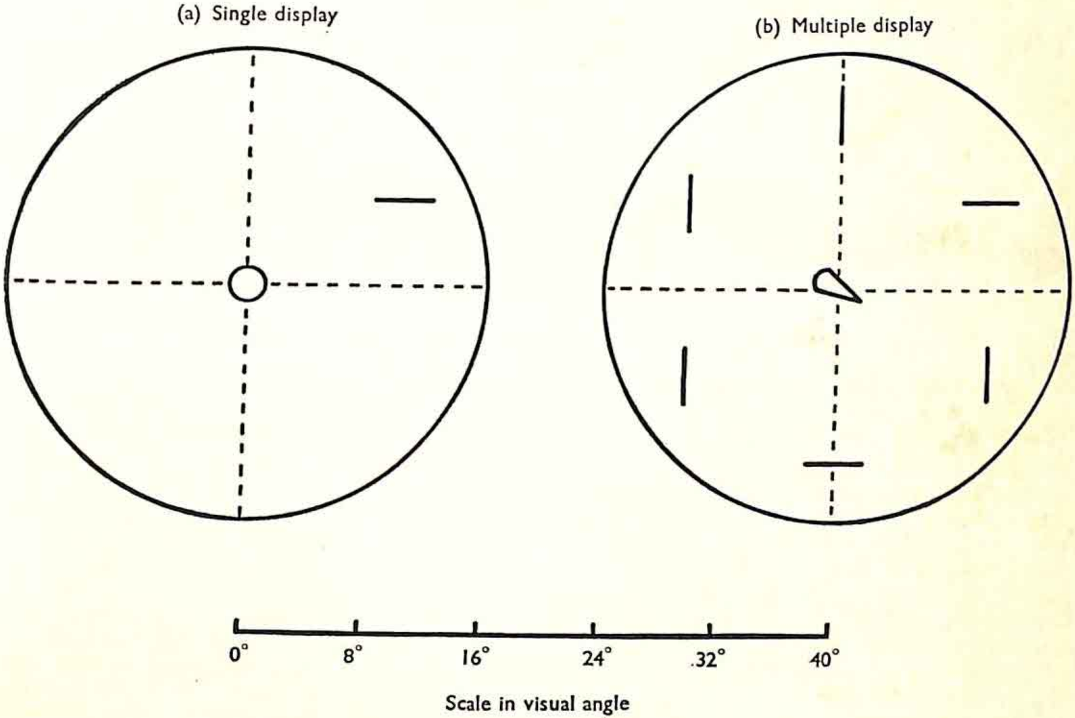
Diagram of apparatus used in Experiments 2 and 3.

Apparatus. The apparatus was a rather complicated tachistoscope employing the principle of Maxwellian view and a diagrammatic vertical section is shown in Figure 2. The subject sat with his head against a rest (RR) and his right eye looked through an eye piece (E) which had an entrance pupil of about 4 mm. diameter. The eye piece embodied a crosswire and was focussed so that the crosswire was at infinity. The object glass (L) was a photographic anastigmatic lens of 3¼ in. focal length of F4·5 aperture. Beyond this was an aluminium disk (DD) mounted on point bearings and weighted with lead so that it had a period of swing of about 1 sec. This disk had a small window (W) which, as it passed the object glass (L), gave the subject an instantaneous view (about 0·05 sec.) of the objects beyond. After a half swing the disk was automatically caught so that the subject did not get repeated exposures. The front of the disk was white and its illumination could be adjusted from Chamber 1. The optical system was adjusted so that the disk was approximately in the critical Maxwellian position, i.e. so that the subject had no impression of movement when the aperture swung across his line of vision. Just beyond the disk was a "beam-splitter" (BS). *Through* this beam-splitter the subject could see clearly in focus any test object on a card in a holder (C1) which was illuminated from Chamber 2. *Reflected* by the beam-splitter and also by a right-angled prism (P) the subject could see another card in holder (C2) clearly in focus. Chamber 3 was not illuminated and the card in C2 was opaque but with a small central window which was used as an attention indicator and through which the subject could see the back of Chamber 4 which was illuminated. Card holder (C3) and Chamber 5 were not used in this experiment.

Display. For simplicity only horizontal and vertical lines were used in the display and the subject was therefore required to say whether a line in a particular part of the visual field was horizontal or vertical. These lines could be shown in any or all of six positions as illustrated in Figures 3*a* and *b*, these positions corresponded to the *even* clock hours (12, 2, 4 . . .). Thus it was possible by changing the cards in the holders C1 and C2 to arrange that the subject had a direction indicator as in Figure 3*b*, or no indicator as in Figure 3*a* for any of a variety of single or multiple displays. The trials in which an arrow was used as an indicator, i.e. where the subject was given instructions simultaneously with the exposure, will be referred to for brevity as K trials. Those in which a circle indicating *no* immediate instruction will be referred to as N trials. Similarly, single displays will be abbreviated to "S," and multiple displays to "M." For example, Figure 3*a* would be referred to as an SN trial, and Figure 3*b* as an MK trial.

Procedure. Examples of four types of display SN, SK, MN and MK, were shown to the subject and he was given a few practice trials with each type of display. In the actual experiment he was given a warning "Ready" which meant he was to get his eye into position so that when he looked at the fixation cross the rest of the field appeared uniform. Then he was told "Fix" and he fixated on the crosswire. At about 1 sec. after this instruction the display was exposed. When the trial was a "K" one (i.e. when there was an arrow at the centre of the field), the subject often reported immediately on what he had seen at the position indicated. In an "N" trial, however, the experimenter waited for 2 sec. after the exposures before naming the clock hour about which a report was wanted. Thus the correct answer in Figure 3*a* would be "Two horizontal," and in Figure 3*b*, "Four vertical."

FIGURE 3



Typical displays used in Experiment 2.

Each session comprised 48 exposures, one vertical and one horizontal in each of the six clock hour positions used under each of the four conditions SN, SK, MN and MK. All 48 trials were randomized and since the same actual displays were used for K and N trials any inaccuracies in their preparation could not cause systematic errors.

In each trial the subject was instructed to give his answer in the form "horizontal," "vertical" or if he had seen the object but felt doubtful about his answer he was asked to add the word "doubtful." There were also some cases when the subject reported he had not seen the test object at all or that he had not the slightest idea which way up it was. In these cases we at first asked for a forced guess. But this procedure became rather farcical as some of the subjects disliked guessing about objects which they said they had not seen at all, and others resorted to such methods as pulling coins from their pocket. Quantitative results therefore take into account only the cases where the subject gave an unqualified answer.

Results. The quantitative results, i.e. the total number of cases in which the subject gave an unqualified and correct answer out of a possible total of 120 were as follows: SK 62, SN 61, MK 56. It is obvious that under these three conditions we were working very close to the subject's threshold for acuity, i.e. his total number of correct and unqualified answers was very near to the number (60 out of a possible 120) to be expected by chance guessing. On the other hand if we consider the fourth condition MN, the position is very different, for here the total of correct and unqualified answers was only 28, again out of a possible 120, which is significantly different by the Wilcoxon test at least at the 0·02 level from each of the other three conditions. The other three conditions compared by the same test showed no significant difference at any level.

There were hardly any cases under any of the four conditions where a subject gave an unqualified answer that proved to be wrong, and in the exceptional cases when this did occur we attributed it to a slip of the tongue or a mistake in what clock hour was intended. When we attempted to take into account the cases in which the subject's answers were qualified by the words "doubtful" or "guess" we found that the data were insufficient to allow any satisfactory statistical treatment. A striking feature of the subjects' reports was that under the MN conditions they frequently said that they *might* have seen something at the position named but could not remember what it was. Sometimes, however, they were able to remember what they had seen in some of the other positions. None of the subjects reported positive or negative after images of any of the peripheral test objects even when questioned on this point.

EXPERIMENT 3. *Differential Threshold with a Simultaneous Attention Signal.*

This experiment was intended to find whether similar laws applied to differential luminance thresholds as those which had been indicated in Experiment 2 for acuity thresholds.

The apparatus and the general experimental design was the same as those in Experiment 2. In this case, however, cards with holes in them were used instead of cards with lines. The illumination of the back (B) of Chamber 5 could be varied by an adjustable slide so that a punched hole could be seen as a spot either dimmer or brighter than the background. The cards were inserted in the holder C3 rather than in C1. This made them slightly out of focus during the exposure so that any imperfections in the edges of the punched holes would not be noticeable to the subject. Thus if the subject had focussed on the cross wire, the attention indicator (which was identical with that used in Experiment 2) would be clearly in focus during an exposure but anything on a card at C3 would be slightly blurred.

In the trials where the display was a single spot as in Figure 4*a* the peripheral spot was made either about 15 per cent. brighter or 15 per cent. dimmer than the background. With *multiple* displays the problem was somewhat more complex and we found it simplest to stick small pieces of cardboard of greater or lower reflectance to the back of four of the six holes. Thus in the multiple display illustrated in Figure 4*b*, the holes at 12 o'clock and 6 o'clock were backed by dark cardboard, while those at 2 o'clock and 10 o'clock were backed with light cardboard. The other two holes—4 and 8 o'clock—could be varied from bright to dark spots in the same way as the spot in the *single* display.

The instructions and procedure were exactly the same as that in Experiment 2 except that the subject was instructed to give his answer in the form "brighter" or "dimmer" (i.e. brighter or dimmer than the background).

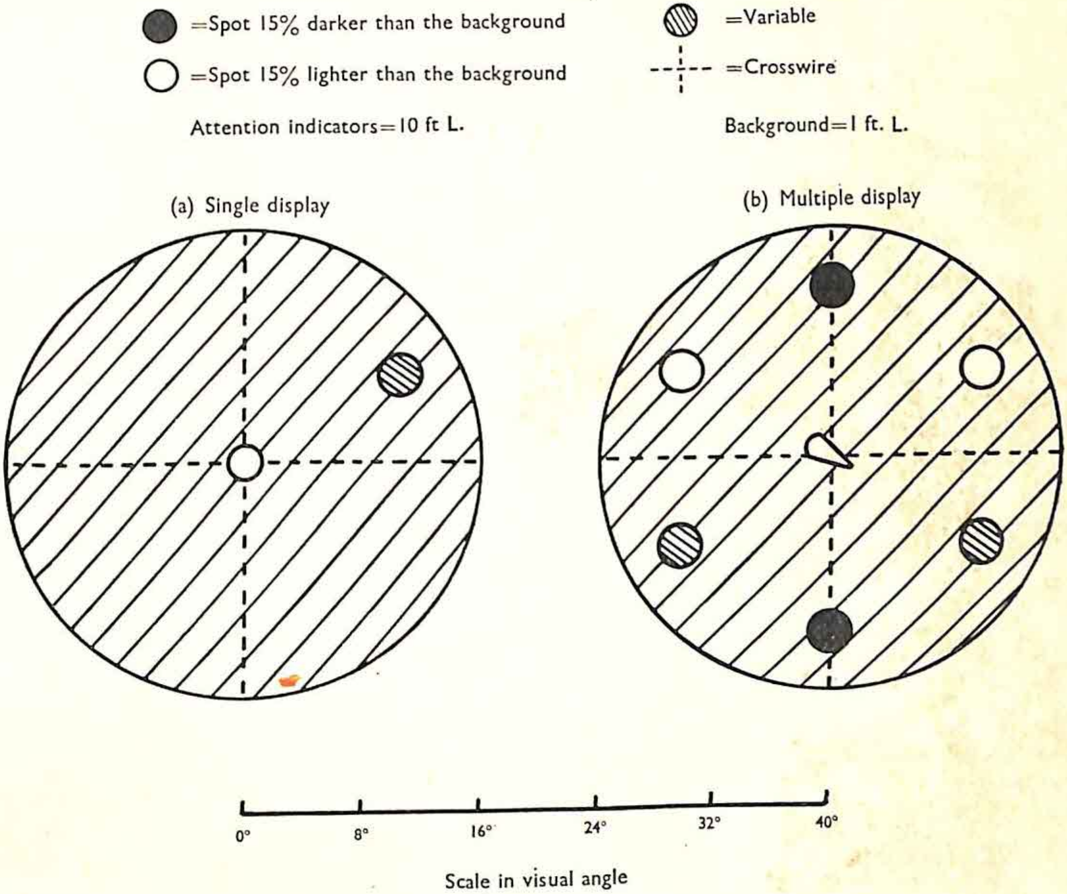
Results. The quantitative results were as follows: out of a possible total of 120 the number of unqualified and correct answers were SK 70, SN 73 and MK 75. Under the fourth condition, MN, there were only 30 correct answers; and this is significantly different by the Wilcoxon test at the 0·01 level from each of the other conditions. As in Experiment 2 the other three conditions showed no significant difference at any level.

No after images of the peripheral objects were reported.

Discussion

Experiments on the effects of attention on peripheral vision have a peculiar difficulty in that the natural tendency when asked to attend to a particular part of the peripheral field is, of course, to turn one's eye in that direction. Nevertheless, as William James (1890) remarked . . . "Practice, however, enables us, *with effort*, to

FIGURE 4



Typical displays used in Experiment 3.

attend to a marginal object whilst keeping the eyes immovable." Our own experiments strongly support this observation by James and (in Experiment I) give some objective confirmation. We found that many subjects initially made eye movements towards the part of the field to which they were asked to attend but after a number of practice trials very few such eye movements were observable, though the subjects reported that they in fact usually attended as directed. In the few cases where detectable eye movements were made, or where the subject reported that he was not attending as directed, that particular reading was cancelled. (It was, however, repeated later in the run; but the subjects were not told that this would happen.)

It is extremely difficult to give a subjective description of what happens when we attend to a peripheral object. Professor C. D. Broad said "I know that when I am attending to an object on the right of the fixation point I am doing *something* different from when I am attending to an object on the left." This was about as far as most of our subjects would go. Several other subjects, including two who described themselves as "hardboiled behaviourists." said they could obey the instructions but could not define or describe the nature of attention. Other subjects, including some who thought it possible that the physiological basis of attention was the sensitization of some part of the retina or visual pathway, were usually able to say whether or not they were attending as instructed during a particular trial. In general we found no evidence that a subject's preconceived views on the nature of attention had any influence on his results. The important point is that all the subjects said that they understood the instructions and were all able to say after (or during) a session whether they, in fact, obeyed the instructions.

The main conclusion from our experiments is that when attention is directed to one part of the peripheral field it has little or no effect on the visibility of a *single* test object (tachistoscopically exposed) in that position. If, however, other objects are exposed simultaneously in other parts of the peripheral field the effect of attention is very striking. This seems to be true even when all the objects concerned are near to threshold value.

In Experiment 1 (where *foreknowledge* of the position of an acuity test object could be given) we had a rough check on eye movements. In view of earlier work on the variation on acuity with eccentric angle (Mendelbaum and Sloan, 1947) it seems unlikely that an eye movement too small for us to detect would have had much effect on our results. It was also pointed out to us by Professor Matthew Alpern that when one is instructed to attend to something there can be involuntary changes in accommodation and pupil size which might affect our results. We had no objective check on these but the fact that with a single test object, the instruction to attend to a particular position had no significant effect on the accuracy, strongly supports the view that none of these changes (eye movements, accommodation or pupillary responses) had much influence on the results.

We did not do a parallel experiment on the effect of *foreknowledge* of the position of the test object on luminance thresholds (because of the obvious technical complexities of arranging a display in which there could be tachistoscopically presented objects either brighter or dimmer than the background, and which would yet allow a control of eye movements). There is, however, an experiment by Mertens (1956) which indicates that in the case of *single* stimuli presented in this way foreknowledge of position had no significance. This is similar to our own results for acuity (in Experiment 1).

In Experiments 2 and 3 where the exposure time was only about 50 millisec. and the attention indicator appeared simultaneously with the test objects, the effects of any changes in the adjustment of the eye can be neglected. Under these conditions the subject had no foreknowledge of what was to occur in any particular trial and no time to make an adjustment of the eye during an exposure. Thus if he was not correctly fixated or accommodated at the moment of any exposure this could only result in an increased scatter of the results with no systematic bias.

It is very difficult from psychological experiments such as those described in this paper to draw any conclusions about the physiological or anatomical basis of attention. Nevertheless the results of Experiment 1 combined with Merten's results make it difficult to believe that the effect of attention to a particular part of the peripheral field is due to any sensitization of the corresponding part of the retina or visual pathway. These results combined with those of Experiments 2 and 3 do, however, suggest that the role of attention in peripheral vision is not in any way to modify the messages sent to the brain from any part of the retina but rather to ensure that when there are competing messages from other parts of the visual field (multiple stimuli) the messages from the part to which attention is directed are given priority of treatment. When there is only a single stimulus, i.e. no competition from other parts of the field, the direction of voluntary attention seems to have no effect. The results of Experiment 2 and 3 also suggest (as do those of Averbach and Sperling, 1960) that this kind of selective attention acts on the immediate memory trace of the stimulation. It therefore seems most profitable to consider the question of attention in peripheral vision as one of the filtering of the available information from a short-term information store as has already been suggested by Averbach and Sperling (1960). This concept of the function of attention is, of course, not new. It has been discussed at greater length and in other contexts than that of peripheral vision e.g. by Broadbent (1958) and Deutsch and Deutsch (1963).

We have in this paper discussed only some of the conditions under which attention to a part of the peripheral field has a "positive effect," i.e. makes perception clearer. There is, however, some evidence, e.g. by Babington Smith (1961*a* and *b*) that *prolonged* attention to a peripheral object causes the reverse effect, i.e. it causes the object to disappear more rapidly than other objects in the field. We have confirmed some of these results and hope to describe our findings in a later paper.

We should like to express our thanks to Professor O. L. Zangwill for encouragement, and to the Medical Research Council for a personal grant to one of us (V.T.) and for an apparatus grant.

REFERENCES

AVERBACH, E., and SPERLING, G. S. (1960). Short-term storage of information in vision. *4th London Symposium on Information Theory*, p. 196.

BABINGTON SMITH, B. (1961*a*). An unexpected effect of attention in peripheral vision. *Nature*, **189**, 776.

BABINGTON SMITH, B. (1961*b*). Effect of attention in peripheral vision. *Nature*, **191**, 732.

BROADBENT, D. E. (1958). *Perception and Communication.* London: Pergamon.

DEUTSCH, J. A., and DEUTSCH, D. (1963). Attention: Some theoretical considerations. *Psychol. Rev.*, **70**, 80–90.

GRINDLEY, G. C., and TOWNSEND, V. (1965). Binocular masking induced by a moving object. *Quart. J. exp. Psychol.*, **17**, 97–109.

GRINDLEY, G. C., and TOWNSEND, V. (1966). Further experiments on movement masking. *Quart. J. exp. Psychol.*, **18**, 319–26.

HELMHOLTZ, H. von (1925). *Handbuch der Physiologischen Optik.* English Translation, Southall, J. P. C., Vol. 3, Chap. 31, p. 455.

JAMES, W. (1890). *Principles of Psychology.* Dover Publications Inc., Vol. 1, p. 437.

MENDELBAUM, J., and SLOAN, L. L. (1947). Peripheral visual acuity. *Amer. J. Ophthal.*, **30**, 581–8.

MERTENS, J. J. (1956). Influence of knowledge of target location upon the probability of observation of peripherally observable test flashes. *J. opt. Soc. Amer.*, **46**, 1069–70.

SPERLING, G. S. (1960). The information available in brief visual presentations. *Psychol. Monogr.*, **74**, No. 11.

Manuscript received 12th July, 1967.

CROSS-MODAL INDUCTION OF CHANGES IN SENSORY THRESHOLDS

BY

GAIL R. HARDY* and D. LEGGE

From Department of Psychology, University College London

Two experiments are reported in which subliminal stimulation of one modality with emotional material impairs detection performance in a different modality. In the first experiment the visual awareness threshold for neutral material was raised by simultaneous auditory presentation of emotional words. This result supports the hypothesis that threshold changes induced by emotional stimulation are mediated centrally. In the second experiment a similar effect was obtained when the experimental roles of the two modalities were reversed. An analysis of the effect in terms of the parameters of Signal Detectability Theory indicates that it is mediated by a lowering of the sensitivity of the detection mechanism. This may be caused either by attenuation of incoming signals, or by an increase in the level of "noise" against which the signal is received.

INTRODUCTION

The term "perceptual defence" may be taken to mean any relation between stimulus emotionality and recognition threshold or awareness threshold, although the term usually refers to a raising of threshold associated with "emotional" stimulation.

In early perceptual defence experiments (e.g. McGinnies, 1949) the subject was required to recognize "neutral" and "emotional" words presented either tachistoscopically, or at very low intensities, and it was found that the recognition threshold was higher for emotional words than for neutral words. This type of design was subject to much valid criticism; for example, that subjects might suppress verbalizing "taboo" words, or that recognition threshold would vary according to the subject's familiarity with words (which, presumably, would be greater for neutral words). Many later experiments adopted more sophisticated techniques (cf. Brown, 1961) which attempted to control for these, and other, factors influencing threshold.

Nevertheless, the criticisms of perceptual defence experiments have been numerous and directed at both methodology (Eriksen, 1954; Singer, 1956; McConnel, Cutler and McNeil, 1958; Klein, 1959) and interpretation (Luchins, 1950; Rosenstock, 1951; Postman, 1951; Lysack, 1954; Putsell, 1957).

In 1961 Dixon and Haider reported the results of an experiment which controlled for word length, word familiarity, and, as far as possible, for letter content, in which it was found that the subliminal presentation of emotional words (i.e. sexual and taboo words) to one eye produced significant changes in the absolute brightness threshold of the other eye. The results of another experiment (Dixon, 1960) suggested that perceptual defence depends upon central as opposed to peripheral mediating mechanisms.

In considering the hypothesis that the activating systems of the brain stem and thalamus might be involved in perceptual regulation, Dixon referred to two relevant facts:

(1) The neural transmission rate over the specific pathways from the sense organs to the cortex, and cortex to reticular system exceeds that over the collaterals from the sense organs to the reticular system.

---

* In receipt of a Science Research Council Postgraduate Studentship.

(2) When the reticular system is depressed by anaesthesia, arrival of visual signals at the cortex may still be detected under conditions which would preclude their conscious reception.

These facts led Dixon (1962) to suggest that the raising of thresholds for emotional stimuli might stem from a lack of cortical arousal resulting from depressed reticular activity following cortical discrimination of the stimulus.

If this is the case, depression of the reticular system should have a general effect in raising thresholds not only of the specific modality stimulated, but of other modalities also.

Experiment 1

The first experiment reported here was designed to test the hypothesis that emotional stimuli presented auditorily below the awareness threshold will produce higher awareness thresholds for neutral visual stimuli than will neutral auditory stimuli.

Method

Visual awareness thresholds for neutral stimuli were measured during concurrent subliminal auditory stimulation with either emotional or neutral words. Subliminality was achieved by recording the stimulus words on one track of a stereophonic tape, whilst "white noise" was recorded on the other track. Independent adjustment of the volumes of words and white noise permitted the auditory stimuli to be presented at approximately the same number of decibels below threshold for each subject.

The visual awareness threshold was measured for two stimuli; the inverted mirror-image of the word "bridge" in small type, and a solid rectangle of light. Visual thresholds were estimated by the Method of Limits, using descending trials only. Auditory presentations began immediately prior to each threshold assessment, and did not cease until the threshold had been recorded for that trial.

Subjects

There were 16 subjects in this experiment, eight male and eight female, all of whom were laboratory staff in the Department of Anatomy, University College London.

Apparatus

The subject sat in a cubicle wearing earphones and facing a small translucent screen. The earphones relayed the auditory stimuli from a stereophonic tape-recorder wired so that each ear-piece received transmission from both tracks. Background illumination of the screen was provided by a projector behind the screen. The visual stimuli were photographic negatives projected by means of a second projector. The intensity of the visual stimuli was controlled by means of a calibrated polaroid filter placed in front of this projector. The filter was operated manually by the experimenter, and its calibrations were used in measuring thresholds.

Selection of auditory stimulus material

Twelve "emotional" and 12 "neutral" words were chosen fairly arbitrarily by the experimenter, and these 24 words were presented orally in random order to 50 non-experimental subjects. These subjects were asked to rate each word for emotionality on a seven-point scale. Subsequently, four words of each type were selected which received consistently high or low emotionality ratings.

The words chosen were:

Emotional (E) cancer, danger, failure, orgasm

Neutral (N) cattle, tanker, letter, dell.

Procedure

An attempt was made to reduce possible adaptation to the auditory stimuli by presenting two words from a given class (i.e. E or N) on each trial, the two words being spoken alternately. Since there were four words in each class, there were six possible pairs of words in each class giving a total of 12 auditory conditions. For each auditory condition there were two possible visual stimuli—the solid rectangle (R) and the word (W), making a total of 24 experimental conditions. Each subject was tested under every condition, performing 24 trials.

Each subject underwent the 24 trials in a different random order. Since the various tape-recorded auditory stimuli were identified by random code numbers, and randomly situated on the tape, the experimenter was totally unaware of the nature of the auditory stimuli being presented to the subject during the experiment. In this respect a "double-blind" procedure obtained. The 24 trials were presented in four blocks of six trials with a 2-min. rest between blocks. Each block was preceded by a reassessment of the subliminality of the auditory stimuli. The alphabet was recorded at the beginning of the tape in the same voice and at the same intensity as the experimental auditory stimuli. Before each block of trials, this piece of tape was played to the subject and the volume of the "voice" track was gradually increased until the subject reported irregularity of the white noise—a phenomenon found to precede detection of the presence of the voice. The "voice" track intensity was then reduced until the irregularity ceased, and auditory presentations in the following block of trials were made at this intensity.

In order to discourage the subject from imagining the presence of a visual stimulus, two dummy trials, on which no visual stimulus was presented, were randomly situated among the experimental trials.

The subject was informed in his instructions that the experiment was concerned with the effects of different noises on concentration, and was told to ignore the noise as far as possible. Two practice trials preceded the experimental trials. The threshold for the visual stimulus was taken as the setting of the polaroid filter when the subject reported the disappearance of the stimulus.

After all experimental trials had been completed, the subject was asked to describe the noise that he had heard during the experiment.

Results and discussion

No subject reported hearing anything other than the white noise, which was frequently described as "a rushing noise."

The mean threshold, in arbitrary units, for each condition, averaged over subjects, is shown in Table I.

TABLE I

AVERAGE VISUAL AWARENESS THRESHOLDS, IN ARBITRARY UNITS, FOR EITHER AN INVERTED, REVERSED WORD OR A SOLID RECTANGLE DURING CONCURRENT EMOTIONAL OR NEUTRAL AUDITORY STIMULATION. (EXPERIMENT 1)

	Auditory stimulation	
Visual stimulus	*Emotional*	*Neutral*
Word	2·00	1·69
Rectangle	3·48	3.04

Almost all subjects showed a rise in threshold for emotional stimuli under both visual conditions. The results of an analysis of variance of these data are shown in Table II.

Both the emotionality of the auditory stimulus and the nature of the visual stimulus produced significant effects on visual thresholds. There was no interaction between these factors.

The large significance of the visual effect reflects the fact that all subjects had higher thresholds for the "rectangle" than they did for the "word," a phenomenon probably related to the fact that the "word" presented a considerably greater length

TABLE II

SUMMARY OF ANALYSIS OF VARIANCE OF THRESHOLD SCORES. (EXPERIMENT 1)

Source	*S.S.*	*d.f.*	*M.S.*	*F*	*Sig. level*
Between subjects	6·95	15			
Within subjects					
Visual	31·72	1	31·72	255·85	0·001
Auditory	2·28	1	2·28	18·39	0·001
Vis. x aud.	0·07	1	0·07	<1	N.S.
Residual	5·57	45	0·12		
Total	46·58	63			

of contour than did the "rectangle." The major concern of this experiment is, however, with the significance of the auditory effect ($p < 0{\cdot}001$).

The awareness thresholds for both types of neutral visual stimulus were higher during subliminal auditory presentation of emotional words than during subliminal auditory presentation of neutral words. The technique of measuring the awareness threshold for non-emotional stimuli avoided some of the problems that beset early experimenters in the field of perceptual defence; however, there are one or two aspects of the method used that require comment.

No attempt was made to match the neutral and emotional words for their frequency of occurrence in language, but this is unlikely to have produced the threshold changes observed, since subjects were never required to recognize these words. The assessment of whether or not the auditory stimuli were subliminal was dependent upon the honesty of the subject during the assessments preceding each block of trials. Whilst this is unfortunate, there is no obvious reason to doubt the honesty of the subjects, particularly since they were not aware that words were available for them to hear.

Whatever importance is attached to these possible artefacts, it seems clear that the results of this experiment were caused by some interaction between the emotionality of subliminal auditory stimuli and visual perception. In view of the cross-modal nature of the effect, it seems likely that some central mechanism is involved; in particular the results are compatible with the suggestion that depression of the reticular activating system is the mechanism responsible for threshold change in response to emotional stimulation.

On consideration of the possible theoretical implications of the results of this experiment, a second experiment was designed to see whether this cross-modal effect could be reproduced when the experimental roles of the two modalities were interchanged.

EXPERIMENT 2

In this second experiment, the emotional and neutral stimuli were presented visually, whilst the subject's task was to detect a faint neutral auditory stimulus. Since this constituted a relatively simple detection task, it seemed appropriate to examine performance in terms of the parameters of Signal Detectability Theory (cf. Tanner and Swets, 1954; Swets, Tanner and Birdsall, 1955; Swets, Tanner and Birdsall, 1961). The two main parameters of this theory are d′, which may be regarded as a measure of the sensitivity of the detecting system, and β, which defines the subject's bias, or criterion, for making particular responses.

In the present experiment, a comparison of changes in d′ with changes in β should enable a distinction to be made between two different central mechanisms which might underlie the phenomenon of perceptual defence. If the value of d′ was lower during emotional stimulation than during neutral stimulation, this would indicate that emotional stimuli raised detection thresholds by somehow reducing the sensitivity of the detecting system. On the other hand, if the subject's criterion (β) was higher during emotional than during neutral stimulation, this would indicate that emotional stimuli raised detection thresholds by somehow decreasing the subject's "willingness" to report the presence of a signal (this does not necessarily imply conscious participation).

The aim of this experiment was to test the hypothesis that the awareness threshold for neutral auditory stimuli will be higher during simultaneous presentation of emotional visual stimuli than during presentation of neutral visual stimuli, when the visual stimuli are presented subliminally.

Should such an effect be detected, the experiment was designed to show whether the effect is due to a change in the "sensitivity" of the perceptual system (d′), or to a change in the subject's bias towards making certain responses (β).

Method

Estimates of subjects' performance in terms of the Signal Detectability Theory parameters, d′ and β, were obtained using the rating scale method (Egan, Schulman and Greenberg, 1959). The subject was required to rate on a four-point scale his confidence that an auditory signal had been presented during a time interval in which subliminal visual presentations of "emotional" or "neutral" words were made.

The auditory signal was a 1,000 cycle per sec. tone, presented on 50 per cent. of the trials, over a continuous background of white noise. The visual stimuli consisted of four "emotional" and four "neutral" words, presented below the awareness threshold.

In order to motivate the subject to look at the screen whilst unaware that words were being presented, the time interval in which the auditory signal might occur was delimited by the presence on the screen of a rectangle of light, and the stimulus word was superimposed on this rectangle. The tone was always presented during the last half second of this 3 sec. presentation, in order to give the subject time to be affected by the presence of the stimulus word. One and a half seconds before this dual presentation of rectangle and word, the rectangle alone was presented for half a second as a warning signal.

Each trial followed the last automatically, and the length of each trial was 18 sec.

Each of the eight stimulus words (four emotional and four neutral) was presented eight times in all, four of these presentations being on trials when a signal was presented, and the other four on "noise only" trials. Since a stimulus word was presented on every trial, there was a total of 64 trials, presented in a different random order to each subject. There was a 5-min. rest-break half way through the sequence.

Selection of visual stimuli

Four words were selected for each category (emotional—E; and neutral—N) in an attempt to counteract any reduction in the effect of the E words on the subjects, which might occur if the same stimulus was presented repeatedly, and also to provide a partial control for individual differences in the perceived emotionality of certain words. The E words were those used in Experiment 1. Previous experiments (e.g. Worthington, 1962) have sometimes used the inverted mirror image of an E stimulus as an N stimulus, and this method controls for many physical factors known to effect the visual awareness threshold. However, this technique reduces not only the "emotionality" of the stimulus, but also its meaningfulness *per se*, and this factor alone might cause a change in threshold to occur. For this reason, four separate N words were selected, matched to the E words as far as possible for length, letter content and structure. The words chosen were:

E: failure, cancer, orgasm, danger

N: feature, fencer, plasma, gander.

A neutral test word, "finger," was also used.

Subjects

There were 14 volunteer subjects in this experiment, seven male and seven female, all of whom were students at University College London.

Apparatus

The subject sat wearing earphones and facing a small translucent screen, with his head supported by a chin rest. The entire area around the subject was enclosed by drapes, to prevent him from inducing peripheral changes in his visual threshold by fixating differentially illuminated parts of the room.

Three projectors provided (*a*) background illumination of the screen, (*b*) the rectangle of light, and (*c*) the word stimuli. These stimuli were in the form of photographic negatives of the words printed in block capitals, their intensity being varied by the use of Ilford Neutral Density Filters. The light intensity of all three projectors was kept constant using a stabilized power unit.

The auditory signal was generated by a Peters Audiometer whilst a separate generator provided white noise. Both white noise and tone were presented binaurally through ear-phones.

The sequence of events on each trial (described in Method) was controlled automatically using nine process timers.

Procedure

The subject was instructed to put on his earphones (through which white noise only was presented), and the "staircase" modification of the Method of Limits (cf. Cornsweet, 1962) was used in conjunction with Spearman's Summation Method to determine the intensity of the test word stimulus corresponding to a probability of detection of 0·3. The "staircase" was continued until 15 reversals of judgement had occurred, after which the inverted mirror images of the four N stimuli were used to ensure that the slight variations in the amount of light transmitted by different words did not produce gross changes in detection threshold. All experimental presentations of stimulus words were made at this intensity.

Whilst a probability of detection of 0·3 may appear to be rather high for a stimulus that is intended to be subliminal, at this stage of the experiment the subjects "undivided" attention was given to detecting the presence of the word, whereas, during the actual experiment, his primary task was to listen for auditory signals.

The intensity of the auditory signal corresponding to a probability of detection of 0·5 was determined, again using the "staircase" method.

The subject was then instructed in the use of the four-point rating scale ("1" representing the greatest confidence that a signal had been presented), and the entire procedure was demonstrated to him, although no stimulus words were presented. The subject made his responses verbally before the end of each trial. The experimental trials were preceded by six practice trials, to enable the subject to practice using the rating scale appropriately.

At the conclusion of the experimental trials, the subject was asked whether or not he thought there had been any words on the screen during the experiment.

Results and discussion

One subject was confident that words had been presented on some trials, and therefore she was replaced. Two subjects expressed doubt as to whether or not words had been presented, but since neither was at all confident about this, their results were included in the analysis.

The data from all subjects were pooled, and the conditional probabilities of reporting a signal when it was present P(Y/S) and of reporting a signal when it was absent P(Y/N) were calculated. A plot of P(Y/S) against P(Y/N) on normal-probability axes indicated that the "signal + noise" distribution had a slightly larger variance than the "noise" distribution. This difference was taken into account by determining both d′ and β in terms of the standard deviation of the "noise" distribution. The plot of the pooled data of all subjects yielded a value of d′ during neutral visual stimulation of 1·73 SD units, and 1·07 SD units during emotional visual stimulation.

When the values of d′ under emotional and neutral visual stimulation were compared for individual subjects, 12 out of 14 subjects showed smaller values of d′ during emotional stimulation ($p < 0{\cdot}01$, using a sign test).

When a confidence rating scale is used, values of β can be assessed at each transition between confidence ratings, each calculated value representing a different degree of caution (Egan, Schulman and Greenberg, 1959). In the present experiment, the calculation of βs for individual subjects was hampered by the fact that subjects did not always use the extreme confidence ratings. This makes it necessary to base calculations of β on estimates of the probability of a response in a category in which no response was made. In view of this difficulty, βs were calculated for the pooled data of all subjects, obviating the necessity for estimation. These values are shown in Table III.

TABLE III

LIKELIHOOD RATIOS FOR THREE IMPLICIT CRITERIA COMPUTED FROM THE FREQUENCIES WITH WHICH EACH OF FOUR CATEGORIES OF CONFIDENCE RATING WERE USED IN AN AUDITORY DETECTION TASK. THE DATA HAVE BEEN ACCUMULATED OVER 14 SUBJECTS (EXPERIMENT 2)

	Criteria		
Concurrent visual stimulation	*1*	*2*	*3*
Neutral	6.78	1·60	0·40
Emotional	3·45	1·37	0·60

The likelihood ratios of the criteria used during emotional stimulation are "telescoped" compared to those used during neutral stimulation. Changes in False Positive Rate (FPR) under the two visual conditions indicate that the positions of decision criteria with respect to the mean of the noise distribution (i.e. x_c) also changed, FPR being highest during emotional stimulation.

The observed reduction in d′ during emotional stimulation indicates a decrease in the sensitivity of the perceptual mechanism due to simultaneous stimulation with emotional visual stimuli.

The observed changes in the likelihood ratios and FPR's of the three criteria are not necessarily due to an effect of concurrent emotional stimulation on the decision-making mechanisms.

The likelihood ratios (βs) obtaining during concurrent emotional visual stimulation suggest a very complex effect of that stimulation, lowering the value of the most cautious criterion and raising the value of the most risky criterion. However, it may be shown that this telescoping effect could result simply from the reduction of d′ provided that the criteria remain at fixed points on the decision axis. By definition, the signal + noise and noise distributions are symmetrically positioned on opposite sides of the point on the decision axis corresponding to $\beta = 1$. A reduction in d′ will be represented by symmetrical movements of these two distributions towards this central datum point. If the subject continues to make decisions on the basis of values on the decision axis, the likelihood ratios of the criteria he uses will change.

The likelihood ratio at a point on the decision axis c SD units from the intersection of the distributions is given by the height of the signal + noise distribution divided by the height of the noise distribution. Since the distributions intersect at a point d′/2 units below the mean of the signal + noise distribution and d′/2 units above the mean of the noise distribution.

$$\beta = e^{-\frac{1}{2}(d'/2 - c)^2}/e^{-\frac{1}{2}(d'/2 + c)^2}$$
$$= e^{d'c}$$

and $\log_e \beta = d'c$.

If d'_0 changes to a new value, d'_1, the new likelihood ratio corresponding to a point c units from the intersection of the distributions is given by

$$\log_e \beta_1 = \frac{d'_1}{d'_0} \log_e \beta_0$$

An estimate of the probability of a false positive response, and hence an estimate of FPR, may be obtained from tables of the normal distribution function after calculating the distance along the decision axis of the criterion from the mean of the noise distribution. This distance, x, is given by

$$x = c + \frac{d'}{2}$$
$$= \frac{\log_e \beta}{d'} + \frac{d'}{2}$$

This procedure was used to predict the likelihood ratio and FPR for each of the three criteria during concurrent emotional visual stimulation and these are shown in Table IV together with the observed values.

TABLE IV

LIKELIHOOD RATIOS AND FALSE POSITIVE RATES OBSERVED DURING CONCURRENT EMOTIONAL VISUAL STIMULATION, AND PREDICTED FROM PERFORMANCE DURING NEUTRAL VISUAL STIMULATION ASSUMING S'S ADOPT A PARTICULAR DECISION STRATEGY

	Likelihood ratios		*False positive rates*	
	Observed	*Predicted*	*Observed (median)*	*Predicted*
1	3·45	3·27	0	0·78
2	1·37	1·34	4	3·21
3	0·60	0·57	8	7·17

The similarity of the observed to the predicted FPRs was assessed for each of the three criteria using a sign test. No significant differences were detected. For the most cautious criterion six of the 14 subjects produced FPRs greater than the predicted value. A similar partition was found for the central criterion. For the least cautious criterion eight of the 14 subjects' FPRs exceeded the predicted value. None of these partitions of subjects are statistically significant. It may be concluded, therefore, that the observed changes in likelihood ratios are adequately accounted for by this model and it is unnecessary to infer any direct effect of emotional visual stimulation on the decision-making criteria.

Although this analysis does not exclude the possibility that the concurrent emotional visual stimulation affected the likelihood ratios of the decision criterion directly, a satisfactory and parsimonious explanation of the changes in these likelihood ratios has been obtained simply by considering the consequences of the reduction in the sensitivity of the detection mechanism (d′).

Discussion

Both experiments indicated a decrease in the sensitivity of one modality to neutral stimuli during simultaneous subliminal stimulation of another modality with "emotional" material. In the first experiment this decrease in sensitivity was detected as a rise in awareness threshold for a neutral visual stimulus during emotional auditory stimulation, and in the second experiment (in which the experimental roles of the modalities were reversed), an analysis of the results in terms of the parameters of Signal Detectability Theory showed a statistically significant decrease in d′ during emotional stimulation.

"Telescoping" of β values occurred during emotional stimulation but this may simply reflect the type of strategy which the subject adopts to re-establish his criteria when d′ decreases. A similar phenomenon has been observed in experiments in which d′ was decreased by dividing the subject's attention between two tasks (Broadbent and Gregory, 1963; Craik, 1965) and it seems possible that a similar explanation to that adopted here may be appropriate. It is tempting to extend the analogy between the two experimental situations, and to suggest that the decrease in d′ during emotional stimulation is caused by division of attention between the two modalities, and that the phenomenon of "perceptual defence" merely reflects the fact that emotional stimuli are more distracting than neutral stimuli. Such a suggestion is untenable, however, because it provides no explanation for the many demonstrations of "unimodal" perceptual defence (cf. Brown, 1961).

The observation made in the second experiment that d′ decreases during "cross-modal" emotional stimulation strongly suggests that perceptual defence is a central rather than peripheral effect, in which the sensitivity of the detection mechanism decreases. The combined techniques of making emotional presentations subliminally, and measuring detection performance for a neutral stimulus presented in a different modality, make it highly unlikely that the observed changes in performance were caused by response suppression, and the nature and extent of the changes in β support this view. It remains possible, of course, that this finding is peculiar to this experimental procedure, in which the subject was deliberately deprived of any obvious basis for making criterion changes: at any rate it is apparent that "perceptual defence" is more than a change in criterion.

The change in the sensitivity of the detection mechanism, implied by the reduction in d′, is compatible with Dixon's suggestion that perceptual defence might result from decreased cortical arousal caused by depressed reticular activity following cortical discrimination of the stimulus. Such a mechanism would indicate a general attenuation of signals reaching the cortex, and, in terms of the parameters of Signal Detectability Theory, would induce a fall in d′ by reducing the general level of cortial activity associated with the "signal + noise" distribution. Another possibility is that the reception of emotional stimuli in some way increases the level of central "noise," thus partially masking incoming signals. Such a mechanism would, theoretically, produce an identical decrease in d′, and it is doubtful whether it is possible to make any distinction between these mechanisms using the usual signal detection procedures. The issue is complicated further by the fact that the effects of an increase in "noise" upon the reception of a signal are likely to vary according to the "absolute" level of noise, and the optimal level of arousal associated with the particular task. A satisfactory answer to these problems must await the results of further experiments.

We should like to thank Professor R. J. Audley, Dr. N. F. Dixon and Dr. T. Shallice for their advice and Mr. J. W. Chambers and his assistants for help with apparatus.

REFERENCES

BROADBENT, D. E., and GREGORY, M. (1963). Vigilance considered as a statistical decision. *Brit. J. Psychol.*, **54**, 309–23.

BROWN, W. P. (1961). Conceptions of perceptual defence. *Brit. J. Psychol., Monogr.* No. 35.

CORNSWEET, T. N. (1962). The staircase method in psychophysics. *Amer. J. Psychol.*, **75**, 485–91.

CRAIK, F. I. M. (1965). *Age Differences in Confidence and Decision Processes.* Unpublished Ph.D. Thesis, University of Liverpool.

DIXON, N. F. (1960). Apparent changes in the visual threshold—central or peripheral? *Brit. J. Psychol.*, **51**, 297–310.

DIXON, N. F. (1962). Feedback and the visual threshold. *J. Communication*, **12**, 97–105.

DIXON, N. F., and HAIDER, M. (1961). Changes in the visual threshold as a function of subception. *Quart. J. exp. Psychol.*, **13**, 229–35.

EGAN, J. P., SCHULMAN, A. L., and GREENBERG, G. Z. (1959). Operating characteristics determined by binary decisions and by ratings. *J. acoust. Soc. Amer.*, **31**, 768–73.

ERIKSEN, C. W. (1954). The case for perceptual defense. *Psychol. Rev.*, **61**, 175–82.

KLEIN, G. S. (1959). On subliminal activation. *J. nerv. ment. Dis.*, **128**, 293–301.

LUCHINS, A. S. (1950). On an approach to social perception. *J. Personal.*, **19**, 64–8.

LYSACK, W. (1954). The effects of punishment upon syllable recognition thresholds. *J. exp. Psychol.*, **47**, 343–50.

McCONNEL, J. V., CUTLER, R. L., and McNEIL, E. B. (1958). Subliminal stimulation: an overview. *Amer. Psychologist*, **13**, 229–42.

McGINNIES, E. (1949). Emotionality and perceptual defense. *Psychol. Rev.*, **56**, 244–51.

POSTMAN, L. (1951). Toward a general theory of cognition. In ROHRER, J. H., and SHERIF, M. (Eds.), *Social Psychology at the Crossroads*, New York: Harper, 242–7.

PUTSELL, T. E. (1957). The experimental induction of perceptual vigilance and defence. *J. Personal.*, **25**, 425–38.

ROSENSTOCK, I. M. (1951). Perceptual aspects of repression. *J. abnorm. soc. Psychol.*, **46**, 304–15.

SINGER, B. R. (1956). An experimental enquiry into the concept of perceptual defense. *Brit. J. Psychol.*, **47**, 298–311.

SWETS, J. A., TANNER, W. P., Jr., and BIRDSALL, T. G. (1955). The evidence for a decision-making theory of visual detection. *Univ. Michigan: Electronic Defense Group, Tech. Rep. No.* 40.

SWETS, J. A., TANNER, W. P., Jr., and BIRDSALL, T. G. (1961). Decision processes in perception. *Psychol. Rev.*, **68**, 301–40.

TANNER, W. P., Jr., and SWETS, J. A. (1954). A decision-making theory of visual detection. *Psychol. Rev.*, **61**, 401–9.

WORTHINGTON, A. G. (1962). *Meaningful Components of a Subliminal Stimulus.* Unpublished Ph.D. thesis, University of London.

Manuscript received 27th July, 1967.

THE EFFECT OF STIMULUS AMPLITUDE ON THE THRESHOLD OF FUSION OF PAIRED FLASHES

BY

P. H. VENABLES and L. A. WARWICK-EVANS

From the Department of Psychology, Birkbeck College, London

Four experiments are described in which the threshold of fusion of paired flashes was measured. The intensity of the single flashes making up the pairs was varied independently. The results showed that when the first flash of the pair was the more intense, the threshold of fusion increased linearly as the difference in the log intensity of the two flashes. A proposal put forward to explain the results suggested the "noise" created by the first flash of the pair has to dissipate to a critical extent to allow the "signal" arising from the second flash to be detected.

INTRODUCTION

In a series of experiments, Horn and Venables (1964) showed that the threshold of fusion of paired light flashes was increased when a shock or a click stimulus was presented in the period from 0 to 400 millisec. before the first of the light flashes and that the amount of increase in the threshold was related to the intensity of this preceding stimulus. A possible explanation for these results is that "noise" is generated by the foregoing click or shock stimuli and that this is added to the disturbance which is created by the first of the flash stimuli. If the reason why a second stimulus cannot be detected immediately after one which precedes it is, that some time has to elapse for the "noise" which is produced by the first stimulus to decay before the second stimulus can be detected, an increase in this "noise" will increase the delay which is required before detection. To test the implications of this model a series of experiments were carried out in which the magnitude of the first and second stimulus of a pair were independently varied. Predictions stemming from the model above could thus be tested. They were (*a*) that the threshold of fusion of paired stimuli should increase as the size of the first stimulus of the pair is increased (*b*) the threshold of fusion should increase as the size of the second stimulus is decreased in magnitude.

While there is little difficulty in directly testing these predictions some ambiguity arises from the statement of the theory on which they are based. Asking the subject to report whether he sees one flash or two flashes, is not strictly the same as asking him to report whether or not he "detects the presence of the second flash." The use of the word "detect" might imply that signal detection procedure should be used and it would follow that a "no-signal" condition should be used in which a single flash is presented. If this is done, however, the single flash is clearly distinguishable from a double flash with any value of inter-flash interval because of the difference in energy content. It is most profitable to view the present experiments in the light of physiological data, such as that of Lindsley (1957), where two flashes at 50 millisec. inter-flash intervals gave rise to only a single evoked potential in the optic cortex in monkey, or where using averaged evoked potential techniques in man (Donchin, Wicke and Lindsley, 1963), suggest that two visual stimuli are only perceived separately when accompanied by average evoked potentials which do not overlap. Similarly, Vaughan and Costa (1964), report that the appearance of two evoked potentials in the visual cortex parallels the subjective experience of perceiving two flashes. Thus in the present context, if a first flash gives rise to an evoked potential response whose size and duration are dependent upon the intensity of the stimulus eliciting them, then it is possible for the evoked response resulting from a small second stimulus close in time to the first stimulus, not to be sufficiently clearly separated from the first evoked potential for the second stimulus to be separately perceived or detected.

METHOD

Pairs of impulses, the interval between which was accurate to the nearest 1·0 millisec. in the case of visual stimuli, and the nearest 0·1 millisec. in the case of auditory stimuli, were produced by a crystal controlled sequence timer ("Digitimer" Devices Ltd.). Pulses from this timer were independently amplified. In the case of visual stimuli pulses 1·0 millisec. in length were applied to the grid of a cathode ray tube (Ferranti CL66). This tube has a phosphor which produces a subjectively white source of light, and has a decay time which ensures that the light flashes are a close approximation in shape to the electrical events which produce them. A circular patch in front of the tube face was viewed binocularly from a distance of 4 ft., the size of the patch resulted in a visual angle of 1·4°. A fixation patch was provided by edge illumination of a clear plastic block placed in front of the tube face. The fixation patch thus occupied the same position as the flash source. The intensity of this fixation patch was 0·02 foot lamberts. The experiments were carried out in a completely darkened room in which the subjects sat for 4 min. before testing began. In the case of auditory stimuli the pulses from the amplifier were each 0·1 millisec. in length and were applied to both earpieces of a pair of moving coil headphones; the amplitude of each pulse could be varied independently. In the description of each experiment the intensity of each stimulus making up a pair is expressed in terms of its log intensity value, in the case of visual stimuli these are given in log foot lambert units, and in the case of auditory stimuli in terms of log voltage across the headphones, a measure which is linearly related to a decibel scale of loudness if a linear response can be assumed for the headphones. In the case of visual stimuli the intensity of the light flashes were measured by a silicon photocell which was itself calibrated by means of a photometer.

The subject was required to press a key once if he perceived one stimulus, or twice if he perceived two stimuli. In no case was a single stimulus ever presented to the subject, examples of "one" stimulus were given by presenting two stimuli with a substantially sub-threshold inter-stimulus interval, and of "two" stimuli by pairs of stimuli separated by long intervals at which all subjects report "twoness." Detection of "twoness" using the cue of brightness was thus minimized. The threshold of fusion was determined by the following technique which was previously described in Horn and Venables (1964). A pair of stimuli with a long inter-stimulus interval (ISI) was presented; if the subject pressed the key twice stimuli with a short ISI were presented; if the subject then pressed once the procedure was repeated with a decreasing range of ISI until two ISI about 1 millisec. apart were reached. Stimuli at these ISI were each repeated twice, and if the response pattern was maintained the ISI at which the subject pressed once was taken as the threshold. The following sequence is from a typical two flash threshold record: 100–2; 50–1; 90–2; 60–1; 70–1; 80–2; 70–1; 75–1; 78–2; 75–1; 76–1; 76–2; 75–1; 75–1; 76–2; 75–1; 76–2. Threshold 75 millisec. As the reliability of such a short procedure might be in doubt an experiment was carried out in which the thresholds of 58 subjects were determined in two periods 4 min. apart. A coefficient of 0·96 was obtained.

Subjects in all experiments were students and staff of Birkbeck College whose ages ranged from 18 to 40 years.

RESULTS

Experiment 1. In this experiment three intensities of flash were used. These were 1·0, 2·0 and 2·5 log foot lamberts. These three intensities of flash were used in pairs in the following six combinations: 2·5–2·5; 2·5–2·0; 2·5–1·0; 2·0–2·5; 1·0–2·5; 2·5–2·5. These combinations of pairs of flashes were presented to twelve subjects in the order of the letters in the rows of two Latin squares, and the threshold of fusion for each pair combination was found. The mean threshold of fusion for each type of pair combination is given in Table I.

TABLE I

THRESHOLDS OF FUSION OF PAIRS OF FLASHES IN SIX INTENSITY COMBINATIONS

EXPERIMENT 1

Intensity combination First flash-second flash Log ft. lamberts ..	2·5–2·5	2·5–2·0	2·5–1·0	2·0–2·5	1·0–2·5	2·5–2·5
Threshold, millisec. ..	76·3	101·5	131·7	84·0	94·8	77·3
Error variance	34·7					

Overall analysis of variance showed that there was a significant difference between these thresholds ($F = 16{\cdot}2$, $p < 0{\cdot}001$, *d.f.* 5/45). Subsequent *t* tests showed that the thresholds for the following intensity pairs were significantly different: 2·5–2·5, and 2·5–2·0 ($p < 0{\cdot}05$); 2·5–2·0 and 2·5–1·0 ($p < 0{\cdot}01$); 2·5–2·5 and 2·5–1·0 ($p < 0{\cdot}001$). The following intensity pairs did not, however, produce significantly different thresholds: 2·5–2·5 and 2·0–2·5; 2·0–2·5 and 1·0–2·5; 2·5–2·5 and 1·0–2·5. This experiment demonstrated that there was a very large increase in threshold when the second stimulus had a smaller intensity than the first. The data suggested that decrease in the intensity of the first flash also raised the threshold, this trend was not, however, significant. It was thought possible that this effect might be due to the decreased visibility of the first flash rather than to its effect in combination with a second more intense flash. The next experiment was therefore performed to investigate this possibility.

Experiment 2. Two levels of flash intensity in four combinations (2·5–2·5; 1·0–1·0; 2·5–1·0; 1·0–2·5) were used to test, first whether a low intensity of the first flash caused a significant rise in threshold and second, whether this rise would be, as expected, dependent on the difficulty of perceiving the first flash alone. If the intensity of first flash alone were responsible then the thresholds for the 1·0–1·0 and 1·0–2·5 conditions would not be expected to be significantly different. Eight subjects took part in this experiment; the combinations of stimuli were presented in the order of the rows of two Latin squares. The mean thresholds found in each condition are shown in Table II.

TABLE II

THRESHOLDS OF FUSION OF PAIRS OF FLASHES IN FOUR INTENSITY COMBINATIONS

EXPERIMENT 2

Intensity combination First flash-second flash Log ft. lamberts	2·5–2·5	2·5–1·0	1·0–1·0	1·0–2·5
Threshold, millisec.	67·2	132·5	81·1	75·2
Error variance	76·0			

Overall analysis of variance showed that the thresholds given to the set of four combinations of intensity differed significantly ($F = 92{\cdot}0$, $p < 0{\cdot}001$, *d.f.* 3/12).

The difference in threshold for the 1·0–1·0 and the 1·0–2·5 conditions was not significant ($t = 0{\cdot}9$); neither was that between the 2·5–2·5 and 1·0–2·5 conditions ($t = 1{\cdot}3$). The small but consistent increase in threshold shown when the first flash has a lower intensity than the second is not significant in either the first or the second experiment due possibly to the use of a sample of insufficient size. The effect is akin to "backward masking," but has not been pursued further in the succeeding experiments.

Experiment 3. The previous experiments showed that as predicted from the model presented in the introduction, the threshold of fusion increased as the intensity of the first flash increased and the threshold increased as the intensity of the second flash decreased. These two statements may be re-expressed as one by saying that the threshold of fusion is increased as the difference in intensity between the two flashes is increased when the first of the two flashes is more intense. A third experiment was carried out with three intensities of flash used in seven combinations of intensity. These combinations of intensities in log foot lamberts were: 2·5–2·5; 2·5–2·0; 2·5–1·5; 2·5–1·0; 2·0–1·5; 2·0–1·0; 1·5–1·0. Fourteen subjects took part in this experiment, and were presented with pairs of stimuli for which the thresholds were determined in the order of the rows of two Latin squares. The mean threshold of fusion for each intensity combination is given in Table III.

TABLE III

THRESHOLDS OF FUSION OF PAIRS OF FLASHES IN SEVEN INTENSITY COMBINATIONS

EXPERIMENT 3

Intensity combination First flash-second flash Log ft. lamberts . .	2·5	2·5–2·0	2·5–1·5	2·5–1·0	2·0–1·5	2·0–1·0	1·5–1·0
Threshold, millisec.	79·8	93·8	106·1	127·9	94·1	111·6	87·9
Error variance	114·5						

Analysis of variance of these data show a significant overall effect of intensity combination ($F = 32{\cdot}1$, $p < 0{\cdot}001$, *d.f.* 6/60). The first four figures in Table III show that there is an almost exactly linear increase in threshold with increase in difference between intensities of first and second flashes. Comparison of the thresholds for the intensity combinations 2·5–2·0; 2·0–1·5; 1·5–1·0 where the difference in intensity between first and second flashes is in each case 0·5 log units, shows that they do not differ significantly (for the difference between the most widely separated of these thresholds $t = 0{\cdot}31$ N.S.). Similarly, a comparison of the thresholds for the intensity combinations 2·5–1·5 and 2·0–1·0 shows in this case where the difference between flash intensities is 1·0 log units that the thresholds do not differ significantly.

Experiment 4. The previous experiment showed that while the difference in intensity was of marked importance in the determination of the threshold of fusion of paired flashes, the intensity of stimulation about which this difference was made did not seem to be significant. A final experiment was therefore carried out with the aim of further establishing the importance of the difference between intensities of the pairs of flashes in the determination of their threshold of fusion, and also establishing the role of intensity as a factor influencing threshold and possibly interacting with intensity difference to influence threshold of fusion. A further variable was introduced into this experiment as a check on the influence of peripheral factors on the fusion threshold. It might be thought that when a first flash is intense it would produce an after-image that was sufficiently marked to interfere with the subjects' perception of the second flash and thus lengthen the threshold. While the results from Experiment 3 mitigate against this sort of explanation insofar as the threshold of fusion is to a large extent similar for intensity differences between flash pairs at different levels of absolute flash intensity a positive check on this factor can be made by the use of another modality. The whole of the design for the visual modality in this experiment was therefore duplicated for the auditory modality. Five intensities of stimulation in nine intensity combinations were used in this experiment. Table IV shows the visual and auditory intensity combinations used and the thresholds obtained in each instance. As before the visual intensities are given in log foot lambert units, while the intensities of auditory stimuli are given in log (voltage × 100).

TABLE IV

THRESHOLDS OF FUSION OF PAIRS OF FLASHES AND PAIRS OF CLICKS IN NINE INTENSITY COMBINATIONS HAVING THREE LEVELS OF DIFFERENCE OF INTENSITY BETWEEN MEMBERS OF PAIRS AND THREE LEVELS OF ABSOLUTE INTENSITY

EXPERIMENT 4

Intensity difference in log units	*1·0*			*0·5*			*0*		
Visual stimuli intensity combinations First flash-second flash Log ft. lamberts ..	3·0–2·0	2·5–1·5	2·0–1·0	2·5–2·0	2·0–1·5	1·5–1·0	2·0–2·0	1·5–1·5	1·0–1·0
Threshold, millisec.	100·0	104·0	118·0	87·0	83·0	85·0	75·8	80·0	81·1
Auditory stimuli intensity combinations First click-second click Log voltage × 100 ..	3·0–2·0	2·5–1·5	2·0–1·0	2·5–2·0	2·0–1·5	1·5–1·0	2·0–2·0	1·5–1·5	1·0–1·0
Threshold, millisec.	4·1	4·1	3·3	2·9	2·7	2·7	1·8	2·0	1·7

Error variances (i) for intensity differences First flash-second flash— 42·42
(ii) for absolute intensities 85·76
(iii) for modalities 1684·10

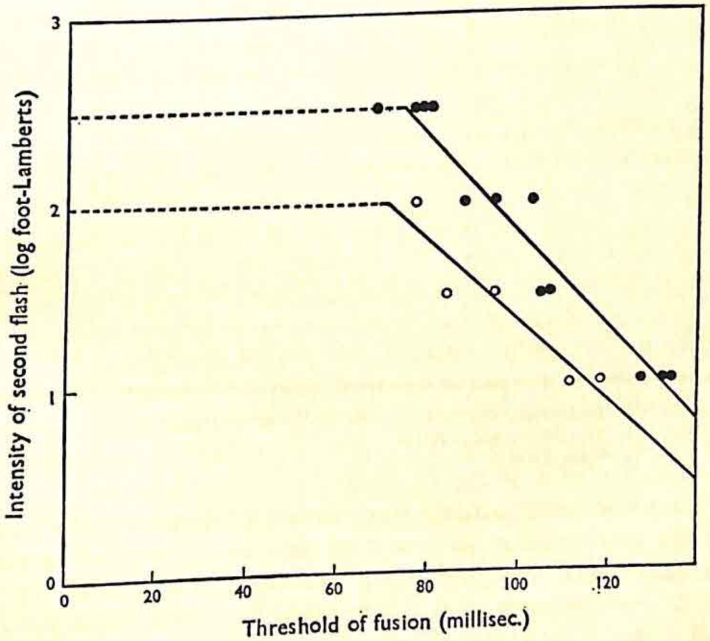
The nine intensity combinations were presented randomly to two groups of six subjects, one group receiving all visual stimuli and the other group all auditory stimuli. The data were analysed using Winer's (1962) model for a three factor experiment with repeated measures. The analysis of variance showed that the effect of difference in stimulus intensity between the first and second stimulus was significant as before ($F = 57{\cdot}8$, $p < 0{\cdot}001$, *d.f.* 2/20). There was also a significant interaction term between intensity difference and modality ($F = 45{\cdot}5$, $p < 0{\cdot}001$, *d.f.* 2/20) showing that the effect of intensity difference was different in the two modalities. There was, of course, a very highly significant term indicating an overall difference in thresholds in the two modalities ($F = 122{\cdot}3$, $p < 0{\cdot}001$, *d.f.* 1/10). The third factor in the analysis was absolute intensity, this was introduced on three levels by taking together the data for the first, fourth and seventh columns, the second, fifth and eighth columns, and the third, sixth and ninth columns in Table IV. In no case was the factor of intensity significant either as a main term or in interaction with any of the other terms. In no case was there any difference in the thresholds to pairs of stimuli having the same difference in intensity between the pairs.

Discussion

The results of all four experiments support the view that the threshold of fusion of paired stimuli increases as the difference in intensity of the stimuli increases where the first stimulus is the more intense. The relation holds over the range of absolute intensities of stimuli used in the experiments described. These results are consonant with the notion put forward in the introduction, that the first stimulus of the pair sets up a neural disturbance or "noise," and that this "noise" has to dissipate before the signal resulting from the second stimulus can be detected.

To illustrate this point more succinctly, and also to take account of the results of all four experiments at the one time, the data shown in Figure I were plotted. The filled circles show the thresholds for conditions where the first stimulus of a pair had an intensity of 2·5 log foot lamberts; the intensity of the second member of the pair being scaled on the ordinate. The open circles show thresholds similarly plotted for the conditions where the first stimulus of the pair had an intensity of 2·0 log foot lamberts. Best fit lines were calculated for each set of data. For the data where the first flash has an intensity of 2·5 log foot lamberts the equation of the line is $y' = -0{\cdot}026x + 4{\cdot}40$, and for the data where the intensity of the first flash is 2·0 log foot lamberts the equation for the line is $y' = -0{\cdot}022x + 3{\cdot}53$. The slope of the lines can be considered to represent the rate of dissipation of the noise resulting from the first stimulus. The close fit of the points to the line (shown by correlations of

FIGURE I



Threshold of fusion for paired light flashes when the first flash has the intensity of 2·5 log foot lamberts (upper line) or 2·0 log foot lamberts (lower line) and the second flash intensity is given on the ordinate. Best fit lines show hypothesised decline of noise arising from first flash as a function of time.

$-0{\cdot}97$ and $-0{\cdot}96$ for the first and second groups of data) over all experiments suggest a linear and consistent decline in noise from about 70 millisec. after the first flash, and that this rate of decline of noise is not significantly different ($t = 0{\cdot}8$) for the first stimuli of two different intensities. The caveat that the decline in noise is apparently linear only after 70 millisec. following the first flash is made for the following reasons. Firstly, because this is the range over which data are available. Secondly, because if the best fit lines are extrapolated back towards the time of presentation of the first flash, the implication is that a second flash of greater intensity than the first would result in a lower threshold than for flashes of equal intensity. This is not borne out by experiments on backward masking or by the data from Experiments 1 and 2. In the absence of other evidence it is most economical to suggest that there is no decline in noise until about 70 millisec., but thereafter the fall-off is linear. The data from Experiment 4 suggest that a similar phenomenon takes place in the auditory system but on a different time scale.

Thus far the results have been discussed in terms of development and dissipation of noise. An alternative mechanism to explain the results might be put forward using the concept of refractoriness. It does not seem profitable at the present time to try to distinguish between these alternative proposals, as it may be that psychological, as distinct from physiological refractoriness may perhaps be explained by invoking a "signal in noise" mechanism.

It has been suggested (Venables, 1963, 1967), that the threshold of fusion of paired flashes or clicks might be used as a measure of the level of cortical activation or arousal. This view receives support in that it has been shown by Kopell, Noble and Silverman (1965), that the threshold of fusion of paired light flashes may be raised by the use of a barbiturate, sodium thiamylal, or lowered by the use of methamphetamine. If the suggestion which has been put forward in this paper, that the factor determining the threshold of fusion is the rate of decay of noise which is generated by the first flash is accepted, then a mechanism whereby arousal operates to alter the rate of decay of noise might be suggested. Evarts (1960), described the changes in the spontaneous and evoked discharge of single units in the visual cortex which are brought about by having the animal preparation in a waking or a sleeping condition. He showed with a few exceptions, that when marked reductions in evoked activity occurred during waking, they were accompanied by reductions in spontaneous activity. Data was presented which showed that in general there was an increase in the ratio of evoked to spontaneous discharge, or in other words an increase in the signal/noise ratio, in the waking as compared to the sleeping animal, although there also tended to be an absolute reduction in the evoked discharge. If arousal can be thought of as bringing about an improvement in the signal/noise ratio it is possible that the results of the present experiment and those which use two flash or two click threshold as indicants of activation may be reconciled. Further work requires to be done in which the "noisiness" of the first signal is changed independently of its intensity, and where the supposed dissipation rate of noise is altered by manipulation of arousal.

Without further experimentation it is not possible to relate the present results to those of Mahneke (1958), or Kietzman and Sutton (1968), where alterations of the two flash threshold have been brought about by alteration in flash duration. While some parallels to the present results may be seen in their work by the mediation of Bloch's law little confidence can be felt in the exercise.

It has already been suggested that some of the results of Experiments 1 and 2, where the second stimulus is brighter than the first, are akin to "backward masking" (Raab, 1963); however, the main results reported, where the second stimulus is weaker than the first, are most likely to be concerned with an entirely different phenomenon. The most widely accepted explanation for backward masking, that the more intense second stimulus has a shorter latency than the first (Crawford, 1947; Baker, 1963; Donchin and Lindsley 1965*a*) clearly cannot be an explanation for the present results where the second stimulus is less intense than the first. It could also be argued that the cue of brightness enhancement could be used as a basis for discriminating "oneness." In view, however, of the findings of Bartley (1939), Donchin and Lindsley (1965*b*), that the inter-flash interval for "brightness enhancement" is in excess of 100 millisec. this explanation seems unlikely when most thresholds for stimuli of equal intensity are below this figure.

This work was supported by a grant from the Medical Research Council.

REFERENCES

BAKER, H. D. (1963). Initial stages of dark and light adaptation. *J. opt. Soc. Amer.*, **53**, 98–103.

BARTLEY, S. H. (1939). Some factors in brightness discrimination. *Psychol. Rev.*, **46**, 337–58.

CRAWFORD, B. H. (1947). Visual adaptation in relation to brief conditioning stimuli. *Proc. Roy. Soc. B.*, **134**, 283–302.

DONCHIN, E. and LINDSLEY, D. B. (1965*a*). Visually evoked response correlates of perceptual masking and enhancement. *EEG clin. Neurophysiol.*, **19**, 325–335.

DONCHIN, E., and LINDSLEY, D. B. (1965*b*). Retroactive brightness enhancement with brief paired flashes of light. *Vision Research*, **5**, 59–70.

DONCHIN, E., WICKE, J. D., and LINDSLEY, D. B. (1963). Cortical evoked potentials and perception of paired flashes. *Science*, **141**, 1285–6.

EVARTS, E. V. (1960). Effects of sleep and waking on spontaneous and evoked discharge of single units in visual cortex. *Fed. Proc.*, **19**, 828–37.

HORN, G., and VENABLES, P. H. (1964). The effect of somaesthetic and acoustic stimuli on the threshold of fusion of paired light flashes in human subjects. *Quart. J. exp. Psychol.*, **16**, 289–96.

KIETZMAN, M. L., and SUTTON, S. (1968). The interpretation of two-pulse measures of temporal resolution in vision. *Vision Research*, in press.

KOPELL, B. S., NOBLE, E. P., and SILVERMAN, J. (1965). The effect of thiamylal and methamphetamine on the two flash threshold. *Life Sci.*, **4**, 2211–4.

LINDSLEY, D. B. (1957). The reticular system and perceptual discrimination. In JASPER, H. H., *et al.* (Eds.) *Reticular Formation in the Brain.* London: Churchill.

MAHNEKE, A. (1958). Foveal discrimination measured by two successive light flashes. *Acta Opthalmologica*, **36**, 3–11.

RAAB, D. H. (1963). Backward masking. *Psychol. Bull.*, **60**, 118–29.

VAUGHAN, H. G., and COSTA, L. D. (1964). Application of evoked potential techniques to behavioural investigation. *Ann. N.Y. Acad. Sci.*, **118**, 71–75.

VENABLES, P. H. (1963). The relationship between level of skin potential and fusion of paired light flashes in schizophrenic and normal subjects. *J. Psychiat. Res.*, **1**, 279–87.

VENABLES, P. H. (1967). The relation of two flash and two click thresholds to withdrawal in paranoid and non-paranoid schizophrenics. *Brit. J. soc. clin. Psychol.*, **6**, 60–2.

WINER, B. J. (1962). *Statistical Principles in Experimental Design.* New York: McGraw-Hill.

Manuscript received 22nd August, 1967.

REFRACTORINESS IN INFORMATION PROCESSING

BY

L. H. SHAFFER

From the Department of Psychology, University of Exeter

Two experiments on refractoriness were carried out in which the warning, first and second signals, S_0, S_1 and S_2 respectively, were all single-valued, and the distributions of random intervals between S_0 and S_1, and S_1 and S_2 were the same. In the first experiment the intervals in a trial were statistically independent: the null hypothesis, that the latencies of the two responses would be similar, was rejected, but the results were also found to agree with no existing alternative hypothesis. In the second experiment the intervals in a trial were conditionally related and the second response was found to be faster than in the first experiment. This is discussed in the context of the issue of serial or parallel processing of information.

PART I—REFRACTORINESS AS TRANSFER

In simple reaction tasks a warning signal, S_0, is followed by a signal, S_1, to which response R_1 is required. In refractoriness tasks a third signal, S_2, requiring response R_2, follows S_1. When S_1 and S_2 are single valued the tasks are made non-trivial by randomizing the intervals between S_0 and S_1, and S_1 and S_2.

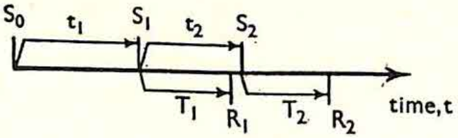
It is known of simple reaction that the latency of response R_1 depends upon the distribution of warning intervals. Two studies particularly relevant here have shown this with intervals between 100 and 900 millisec. Nickerson (1965) found that if all intervals are equally likely then response latency decreases as the interval increases. Also latency is affected both by the range of intervals and by the values spanned, hence he concludes that it is a function of both relative and absolute values of the interval. Karlin (1966) found that if the modal interval is far more probable than any of the others then latency has a minimum value in the region of the mode.

A recent survey of experiments on refractoriness is given by Bertelson (1966), who also traces the history of this research. Interest in these studies has concentrated on the phenomenon that latency of the second response is a decreasing function of the interval between S_1 and S_2. An obvious question is whether this phenomenon is equivalent to the one observed in simple reaction. Yet, surprisingly, although this has been put forward as hypothesis the direct test has not been carried out. If we suppose that S_0 is the warning signal for S_1 and S_1 the warning signal for S_2 then the direct test is achieved by a task symmetry such that

(*a*) the intervals between S_0 and S_1 and between S_1 and S_2 are independent, and

(*b*) they have the same distribution.

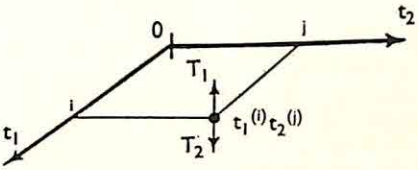
FIGURE 1



The temporal order of events in the refractoriness task is shown in Figure 1, but note that whether or not R_1 precedes S_2 depends upon the timing of S_2 and the latency of R_1. In this diagram t_1 is the interval between S_0 and S_1, t_2 the interval between

S_1 and S_2, T_1 the latency of R_1 and T_2 the latency of R_2. If S_0, S_1 and S_2 are all single valued then the stimulus in each trial can be characterized as a pair of time intervals, t_1t_2, which designates a point in a stimulus plane. This is shown in Figure 2. The interval t_1 takes values $t_1^{(I)}$ at random and t_2 takes values $t_2^{(J)}$ at random.

FIGURE 2



Some fairly general assumptions about performance are that T_1 is a function of t_1 and t_2:

$$T_1 = f\ (t_1;\ t_2) \tag{1}$$

and that T_2 is a function of t_1, t_2 and T_1:

$$T_2 = g\ (t_1;\ t_2;\ T_1) \tag{2}$$

Interestingly, it may not be feasible to test equation (2) in its general form within a single experiment since this would entail sampling each point $t_1^{(I)}\ t_2^{(J)}$ in the stimulus plane a sufficient number of times to be able to estimate the effect of T_1 at every point.

The null hypothesis is that T_1 depends only upon t_1, T_2 depends only upon t_2 and that, if t_1 and t_2 have the same distribution and are independent then T_1 and T_2 have equivalent functions:

$$f'\ (t_1) = g'\ (t_2) \tag{3}$$

Rejection of the null hypothesis implies that there is transfer from one stimulus-response event in the series to the next.

The following experiment provides a test of the null hypothesis, and in case it should fail the test the experimental design was such as to permit the testing of at least some alternative hypotheses. As we have said, no single experiment can provide a strong test of all the hypotheses implicit in equation (2). Any evidence of transfer will be examined in the context of the major current hypotheses of refractoriness.

METHOD

Apparatus. The display consisted of three digital units mounted side by side, each of which displayed a symbol about 1 in. high when illuminated. The left-most symbol (S_0) was an arrowhead and each of the others (S_1 and S_2) was a O. These came on in a left-right sequence and two delay timers were used to introduce a variable delay between onset of the first and second and between the second and third signals (t_1 and t_2 respectively).

The response unit had two keys, each requiring a pressure of about 10 grams to depress, with a traverse of about 2 mm. Two Venner millisecond timers were used, the first measuring the time between onset of S_1 and pressing the left key (T_1), the second measuring the time between onset of S_2 and pressing the right key (T_2). Pressing the right key also terminated the display.

Instruction. The subject was told that the arrowhead was a warning signal, the first O was to be answered by pressing the left key with his left forefinger and the second O by pressing the right key with his right forefinger. He was to answer each O signal as quickly as possible when it arrived and try not to make premature responses. He was told that the intervals between S_0 and S_1 and between S_1 and S_2 would be random.

Procedure. Following instruction each subject was given about 5 min. practice, in which time a variety of intervals between 50 and 310 millisec. were chosen at random for the first and second intervals in each trial. Then there were eight test sessions given two on each of four successive days, one in the morning and one in the afternoon. In each session a different sequence of 125 trials was presented at intervals of about 7 sec. Each sequence used five intervals, 50, 100, 160, 230 and 310-millisec., in t_1 and t_2 and each of the 25 possible combinations of t_1 and t_2 occurred five times. Thus each sequence was made up of trials chosen by random selection without replacement from the 125 possibilities.

Subjects. Three first-year psychology students who were not previously familiar with this type of experiment were used. They were about 18 years old, two men and one woman.

Results

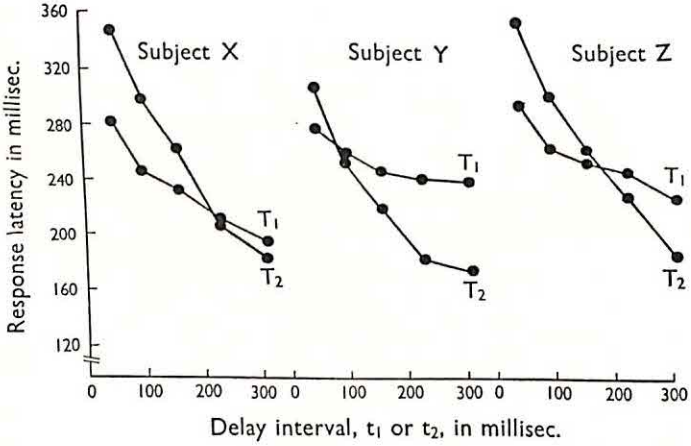
We shall label the subjects X, Y and Z, calling the woman X. The results of each were analysed separately and only subsequently combined.

A further point of procedure that can more relevantly be mentioned here is that if a subject anticipated a signal completely the trial was later repeated. The occurrence of premature responses was not recorded on the first four sessions, but from the fifth onwards it was recorded against the trial on which it occurred. A response was considered premature only if the timer failed to give a reading, i.e. if the latency was less than 1 millisec. These premature responses, whether R_1 or R_2, occurred only at the longest or next-but-one longest interval in t_1 or t_2 and their frequency is reported later.

For each subject the latencies, T_1 and T_2, in a session were separately summed at each point, t_1t_2. Analyses of variance were then carried out examining the effects of t_1, t_2 and sessions, first for T_1, and then for T_2. In case the data did not satisfy all the statistical requirements of the F test, the null hypothesis was considered rejected only if a probability $p \leqslant 0{\cdot}01$ was obtained.

Analysis of T_1. Two effects were highly significant, t_1 and sessions: for t_1, $F\ (4, 112) = 41$ for X, 13 for Y and 71 for Z, $p < 0{\cdot}001$; for sessions, $F\ (7, 112) = 5{\cdot}9$ for X, 7·2 for Y and 18 for Z, $p < 0{\cdot}001$. The effect of sessions did not reflect faster responding with practice, except perhaps in the case of Y. Subject X, if anything got slower with practice and Z was slowest in the middle sessions.

FIGURE 3



Latencies, T_1 and T_2, of the first and second responses as a function of the respective intervals, t_1 and t_2. Part I.

Only one other effect was significant and this was secondary in magnitude. For subject Z the $t_1 \times$ sessions interaction was significant, F $(28, 112) = 2{\cdot}5$, $p < 0{\cdot}01$. This could be interpreted as a decrease in the t_1 effect with practice.

Despite this secondary effect it is fairly reasonable to display the effect of t_1 on T_1 by a single curve for each subject and these are shown in Figure 3. We merely note here that they confirm that response latency decreases as the warning interval gets longer.

Analysis of T_2. Two effects were highly significant, t_2 and sessions: for t_2, F $(4, 112) = 156$ for X, 88 for Y and 226 for Z, $p < 0{\cdot}001$; for sessions, F $(7, 112) = 14$ for X, 10 for Y and 31 for Z, $p < 0{\cdot}001$. The sessions effect is interpreted as for T_1.

There were some secondary effects. For subject Z the effect of t_1 was significant, F $(4, 112) = 5{\cdot}0$, $p < 0{\cdot}01$, and the effect of the $t_1 \times t_2$ interaction was significant, F $(16, 112) = 2{\cdot}6$, $p < 0{\cdot}01$. The first of these reflected an increase in T_2 with the longer t_1 intervals; the second had a complicated form that will be discussed in the combined analysis.

The effect of t_2 on T_2 for each subject is also displayed in Figure 3. There is a striking difference between these curves and those relating t_1 and T_1. As a first approximation it may be said that the T_2 curve is a linear transformation of the T_1 curve, such that there is a point in the delay, t, below which $T_2 > T_1$ and above which $T_1 > T_2$.

Premature responses. As already mentioned these were recorded in only the last four sessions. They were observed to occur only at the longest intervals of t_1 or t_2 (see also Nickerson, 1965). Subject X made 26 premature responses in R_1 and 6 in R_2, Y made 3 in R_1 and 18 in R_2, Z made none in R_1 and 1 in R_2. Taken as an index of a tendency to make anticipatory responses they can be used to provide estimates of how often relatively fast responses appeared in the data. In this case they help explain why the crossover point referred to above occurred fairly late for X and fairly early for Y.

TABLE I

LATENCY IN MILLISEC. AS A FUNCTION OF t_1 AND t_2 AVERAGED OVER SESSIONS AND SUBJECTS. PART I

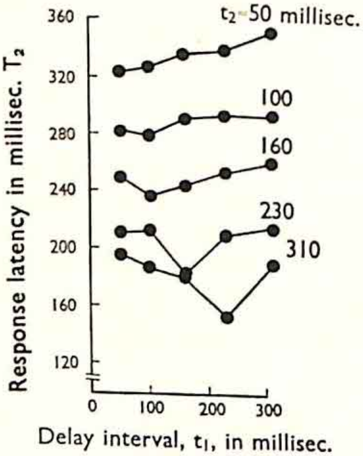
		t_2 *in millisec.*				
T_1		50	100	160	230	310
t_1 in millisec. ..	50	277	278	287	291	289
	100	247	256	248	262	260
	160	240	246	246	236	249
	230	222	234	237	240	229
	310	218	213	219	227	225
		t_2 *in millisec.*				
T_2		50	100	160	230	310
t_1 in millisec. ..	50	324	281	249	210	195
	100	327	279	236	211	186
	160	336	290	243	182	179
	230	339	293	252	208	152
	310	351	292	260	214	189

Combined analysis. Since the results for all subjects were essentially similar two more analyses of variance were carried out on the combined data, for T_1 and T_2 respectively. The data were collapsed over sessions so that only the factors t_1 and t_2 were distinguished. These averaged latencies are shown in Table I.

The analysis of T_1 added no new information; the analysis of T_2 showed a new effect, a $t_1 \times t_2$ interaction, F (16, 32) $= 2{\cdot}8$, $p = 0{\cdot}01$. This was previously observed to be significant only for Z, although a similar pattern was observed in the data of X and Y. The nature of the interaction is clearly shown in Figure 4. At the smallest value of t_2, T_2 increases monotonically with t_1; for increasing values of t_2, T_2 tends to increase with t_1 but has an acute dip that simultaneously becomes more pronounced and moves to larger values of t_1.

No attempt was made to compare statistically the functions $T_1 = f'(t_1)$ and $T_2 = g'(t_2)$, because there seems to be no simple and natural way to allow for the dependence in each yoked pair of observations (T_1, T_2). However, the difference between f′ and g′ is large and demonstrably invariant over subject sessions.

FIGURE 4



Composite latency T_2 as a function of t_1 with t_2 as parameter. Part I.

DISCUSSION

It is clear from the results that the null hypothesis, stating that latencies T_1 and T_2 are symmetric functions of t_1 and t_2 respectively, must be discarded. The curves in Figure 3 show that the functions $T_1 = f'(t_1)$ and $T_2 = g'(t_2)$ are different and the curves in Figure 4 show that T_2 is a function of at least t_1 and t_2. We may therefore proceed to examine whether any current alternative hypotheses offer a transfer mechanism that would produce these results. They must also, of course, account for the general forms of f′ and g′, but as Smith (1967) points out in her review of these hypotheses this is a minimum requirement since these hypotheses were initially presented for just this purpose. The hypotheses fall into one or other of two types, which we may call delay and expectancy.

Delay hypotheses. These postulate that on arrival the signal S_1 preempts and occupies a processing system so that S_2 may have to queue for acceptance. They then divide such that (1) following the acceptance of S_1 there is an interval before acceptance of S_2 that is fixed (Telford, 1931) or has a cyclic reset (Broadbent, 1958), or (2) S_2 must wait until a certain stage of completion of the process relevant to R_1 has been reached (Welford, 1952).

The notion of a fixed delay following acceptance of S_1 provides the origin of the name "refractory period." It postulates a time interval $t_2^{(c)}$ and a response latency, T_a, which has the distribution of an undelayed response, such that,

$$T_2 = T_a \qquad \text{if } t_2^{(j)} \geqslant t_2^{(c)}$$
$$\text{and } T_2 = T_a + (t_2^{(c)} - t_2^{(j)}) \text{ if } t_2^{(j)} < t_2^{(c)}$$

Broadbent's variant of this notion postulates periodic sampling intervals of length $t_2^{(c)}$ such that a signal arriving within an interval is admitted for processing at the end of that interval. Two consequences are that signals arriving in the same interval will be grouped together in processing and that $f' = g'$, which we have already rejected. Alternatively, a further postulate may be added, which synchronizes the sampling cycle with the arrival of S_1, so that S_1 never undergoes a delay. In this case the prediction is the same as for the refractory hypothesis except that the delay interval is recurrent with a period of length $t_2^{(c)}$.

Since the refractory hypothesis postulates a response delay numerically equal to the duration of an unelapsed waiting time it fails the minimum requirement of describing the functions f′ and g′. This could be partly overcome by weakening the hypothesis and making $t_2^{(c)}$ a random variable and so obtaining a better fit of either f′ or g′, but not both. There is no provision in the hypothesis to account for different forms of refractoriness following S_0 and S_1, and there seems to be no natural way to include it.

The second hypothesis, known as the single channel hypothesis, postulates that any signal S initiates a process of duration T_p leading to response R and that processes on different signals follow consecutively, unless the signals are grouped. Since the measured latency, T, of a response is the interval between onset of S and the closing of a switch that stops the timer one must specify a relationship between T_p and T. The simplest assumptions are that they are linearly related. Hence, without grouping,

$$T_2 = T_a \qquad \text{if } T_{1p} < t_2^{(j)}$$
$$\text{and } T_2 = T_a + (T_{1p} - t_2^{(j)}) \text{ if } T_{1p} \geqslant t_2^{(j)}$$

where T_a is an undelayed response latency, as before, and T_{1p} is the latency of the process initiated by S_1. Thus for any interval, $t_2^{(j)}$ there should be a regression of T_2 on T_1, which is a discontinuous straight line having a slope of zero to the left of a point $t_2^{(j)} = T_{1p}$, and a slope of b to the right if $T_{1p} = C + bT_1$.

The transfer mechanism here is the occupancy time of the previous process in the central system. If it is assumed that S_0, as well as S_1 and S_2, can initiate a process then the obvious basis on which to account for the different functions f′ and g′ is that, unlike S_1 and S_2, the process on S_0 does not terminate in an overt response and so should occupy the system for less time. This does not, however, explain why f′ should have a smaller slope than g′, or why these functions should cross over rather than have a common asymptote.

If, however, it is assumed that the effect of the warning interval is other than that of a central delay, then the test of the single channel hypothesis must be based upon the difference function, f′–g′. Not only does this function pass through zero but also it gives delays at small warning intervals altogether too short according to the hypothesis.

The hypothesis predicts the form of regression of T_2 onto T_1 It is difficult to examine this prediction because of the postulated discontinuity in the regression curve. However, at $t_2^{(j)} = 50$ millisec. there are very few values of T_1 less than this and most are considerably higher. It should therefore be safe to assume that the regression curve here has a uniform slope. For each subject the product-moment correlation, r and the regression slope, b of T_2 onto T_1 were calculated at each level of t_1. The results are shown in Table II. The correlations are fairly high and nearly always significant at $p < 0{\cdot}01$. As will be shown later there are reasons to suppose that the positive correlation may be an artefact of instruction. The values of b are usually much less than unity, showing that T_{1p} can be only a fraction of T_1.

TABLE II

CORRELATION, r AND SLOPE OF REGRESSION, b OF T_2 ON T_1 AT $t_2 = 50$ MILLISEC.

		Subject X		Subject Y		Subject Z	
		r	*b*	*r*	*b*	*r*	*b*
	50	0·61	0·66	0·50	0·67	0·66	0·90
	100	0·65	0·68	0·55	0·58	0·75	1·12
t_1 in millisec.	160	0·76	0·70	0·75	0·63	0·51	0·56
	230	0·69	0·40	0·76	0·76	0·17	0·19
	310	0·47	0·44	0·37	0·45	0·34	0·46

$n = 40$ in each test.
$p \leqslant 0{\cdot}01$ if $r \geqslant 0{\cdot}40$.

Expectancy hypotheses. These entail assuming two types of process, although in some formulations one or other of these may not be made explicit (Smith, 1967). One must first assume that there can be different states of preparation to initiate a response given a signal, that there is an ordered relationship between these states and response latency, and that states of greater preparation require a finite time to achieve and can be only transiently maintained (cf. Poulton, 1950). The other assumption is that if the subject knows the distribution of warning intervals then he can use this to predict, or "expect," a point in time following the warning signal at which the signal will arrive and project his state of maximum preparation to this point (Mowrer, 1940).

Such hypotheses can be made precise by specifying (*a*) how latency is related to state of preparation, (*b*) the duration of transient states, (*c*) which properties of the distribution of warning intervals are utilized in prediction, and (*d*) whether and how prediction depends upon recent events in the series.

An additional point to be made is that typically the instruction to the subject in a reaction task provides an ambiguous payoff between responding quickly and making errors or premature responses (Edwards, 1961). He must therefore select his own payoff and combine this with his knowledge of the warning intervals to form a prediction of signal arrival. Annett (1966) suggests that if all intervals are equally likely and the instruction against making premature responses is only moderately punitive then the subject may choose to project his state of maximum preparation to a point in time in the region of the longest interval.

Given this last assumption about payoff an expectancy hypothesis would predict the general features of Nickerson's results on simple reaction (1965) and the forms of f′ and g′ in the present results. It would also explain, as Nickerson pointed out, why premature responses should tend to occur only at the longer warning intervals, as in both his and the present data. Other confirmation is provided in Karlin's results, using leptokurtic distributions of warning interval (1966), showing that fastest responses and most premature responses tend to occur at the modal interval.

However, the current versions of expectancy hypothesis would all predict the equivalence of f′ and g′, since no one has included an explicit statement on sequential modification of expectancy. There is in principle no difficulty in doing so, and a natural way to explain the interaction effect in Figure 4 is to suppose that the subject expects a second warning interval symmetric with the first in that trial.

Commentary. We now seem to have not one effect but two and it is not clear which of these is to be considered refractoriness. There is the effect of warning interval and the transfer effect when one stimulus-response event follows another. The failure to separate these in earlier studies has led to inappropriate tests providing spurious support for one or another hypothesis. Often, these studies have chosen values of t_1 nearly optimal for the first response and so have obtained an exaggerated measure of the transfer effect, tending to favour a single channel hypothesis (e.g. Davis, 1956). Although Welford (1967) considers this hypothesis the most robust in accounting for refractoriness it fares no better than the others with the present data.

Looking back at the general equations for T_1 and T_2 in the introduction can we place any useful restrictions on these in the light of our results? In equation (1) it would seem that t_2 is not a significant variable affecting T_1 and can be omitted. However the possibility arises that it should be replaced by a symbol s, standing for sequence, to indicate that R_1 is influenced by the fact that other events are to follow (cf. Poulton, 1950):

$$T_1 = f\ (t_1;\ s) \tag{1a}$$

This may account for T_1 and T_2 not having the same asymptotic value, and suggests that the experiment presented is incomplete without control data of single reaction time for the subjects ("We know the sound of two hands clapping. But what is the sound of one hand clapping"—a Zen koan). Such data was in fact obtained in a single session with each subject a month after the main experiment. One subject, Y was uniformly faster on the single response than on the first response, the other two showed no difference. This is inconclusive, since one should extensively train the subjects to switch from one task to the other so as to eliminate the carry-over of inappropriate responding habits. Other experiments have confirmed that single response is faster than the first response (e.g. Gottsdanker, Broadbent and Van Sant, 1963). The effect s may be a form of anticipation, as Poulton and Gottsdanker *et al.* suppose, such that the first response is included within a programme for a response sequence initiated by S_0.

In equation (2) there are now grounds for retaining t_1 because it interacts with t_2 in influencing T_2. There are also grounds for retaining T_1 because of the positive correlation with T_2. Such correlation has been taken as support for the single channel hypothesis (Davis, 1956), but it would also be expected on a quite different basis.

The instruction to the subject in a refractoriness task should properly include a statement, "Answer each signal as soon as it arrives." It is perfectly reasonable for the subject to interpret this as implying that R_1 should precede, or at least not follow R_2, since S_1 always precedes S_2. If so then this has two important consequences for response latency:

(1) however late R_1 may occur on a given trial R_2 will only be made at that time or later and, particularly when t_2 is small, this will contribute to a positive correlation between T_1 and T_2, and

(2) at least when t_2 is small, there will be a lower bound on the values of T_2 imposed by the values of T_1, so that the left-hand side of the distribution of T_2 will be curtailed and hence on average T_2 will be larger than T_1. Literally, part of the transfer effect observed may be the effect of how the subject interprets his instruction.

The reason for including this part of the instruction is that otherwise the subject may tend to wait until both signals arrive and make a grouped response. However, with appropriate instruction, it seems a curious procedure to analyse the data, as Welford advocates (1967), by treating separately as grouped responses those cases in which R_2 coincides with or precedes R_1. If the responses were truly independent then one should expect this to occur with a probability calculable by separating the means of two latency distributions by the appropriate value of t_2. It is the virtual absence of simultaneous or reversed responding in the present data that testifies to the non-independence of the two responses.

At present there is justification for an interference hypothesis of serial choice having the following form: any signal relevant to a task may initiate a central process; a process may interfere with any contiguous process so as to prolong it; the amount of interference will vary inversely with their separation in time and with their predictability, separately and in relation to each other. This hypothesis can account in general terms for much of the present results and most of those in the literature (see Bertelson, 1966; Smith, 1967; Welford, 1967). It would be of little worth at the present time to make a stronger hypothesis than this, since what is required is proper parametric data of interference phenomena in serial response. Finally, it is noted that there are already results in serial performance to indicate that the above formulation is inadequate and needs to be supplemented by statements about syntactic properties of the processes initiated on each signal (Bertelson, 1961; Jersild, 1927; Shaffer, 1967).

Part II—Refractoriness and Prediction

In information theory a code is redundant if the symbols in its alphabet are not equiprobable or if there is sequential constraint upon their choice. To the extent that it is redundant the code is predictable. If in a serial reaction task there is sequential constraint between successive signals can the subject utilize this so as to improve his speed of reaction?

The answer to this question is relevant to the issue of whether decisions about stimuli must be made serially or may overlap in time. The former assumption, of course, is the single channel hypothesis (Welford, 1952). Broadbent (1958) has modified this hypothesis to take account of the capacity of the channel, such that stimuli with small uncertainty may be processed together, but this introduces formidable assumptions about the possibility of improvising input codes for composite, unrelated stimuli. These assumptions have never been explicitly developed and fortunately it is unnecessary for our purposes to discuss them further. In the task we shall consider there are observable delays in responding, which are supposed to provide evidence of queuing in a single channel.

The following experiment uses the same task as in Part I, but the warning interval of the second stimulus is now related to the warning interval of the first, providing some opportunity for prediction.

To make clear what prediction must entail we shall analyse the situation in some detail. Let us call the warning, first and second signals S_0, S_1 and S_2, as before, and conceptualize a finite state source emitting these in sequence, with delay intervals t_1 and t_2. The source starts each sequence in a state, A, in which it emits S_0 and goes to one of a number of B states. In the state B_i it emits S_1 after a delay $t_1^{(i)}$ and goes to one of a number of C states. In the state C_j it emits S_2 after a delay $t_2^{(j)}$ and this terminates the sequence. The second delay is related to the first if, for any state C_j, the probability of C_j depends upon which state in B the source passes through. In this case it can be shown that the average uncertainty of t_2 is less than if t_1 and t_2 were independent.

The complete basis for predicting t_1 and t_2 would be the probabilities of transition from one state of the source to another. A less complete basis, one ignoring the sequential constraint, would be the probabilities of each state. It may be, of course, that the subtlety of prediction that can be implemented in response preparation requires an even simpler basis. The response may be organized in relation to, say, the most frequent time interval (Karlin, 1966), taking or not taking into account the sequential constraint.

The knowledge the subject must sustain in order to make any level of prediction can be regarded as a model of the stimulus source, and what matters here is that prediction that utilizes sequential constraint requires a more complicated model than prediction that ignores it. Reduction in uncertainty of t_2 is obtained at the cost of increasing the amount of analysis of the stimulus source. Having observed the value of t_1 in a particular trial the subject must choose one of a set of probability vectors in order to make the best prediction of the value of t_2. Thus when S_1 arrives (and hence terminates the first interval) it seems that the subject must initiate two processes, one to make the first response and the other to predict the second interval. If in fact the latency of the second response is reduced by the sequential constraint then we must ask whether the two processes could occur in series.

METHOD

The apparatus was the same as in Part I. Instruction to the subject was also unchanged.

Procedure. Each subject came for five sessions and in each session he was given a different sequence of 100 trials. The first session was a practice session and the results of this were not used subsequently.

The design of the trials was as follows: t_1 could take any one of five values, 50, 100, 160, 230 or 310 millisec.; t_2 could take any one of these values subject to the constraint given in Table III, which shows how often in 100 trials each value of t_2 followed a given value of t_1. Each sequence was made up by a procedure of random selection without replacement.

TABLE III

		t_2 *in millisec.*				
		50	100	160	230	310
t_1 in millisec. ..	50	10				10
	100		10		10	
	160			20		
	230		10		10	
	310	10				10

Subjects. Four first-year psychology students, who were not previously familiar with this type of experiment, were used. There were two men and two women.

RESULTS

The data for each subject were averaged over the four sessions, having ascertained that there were no marked trends between sessions. Although there were individual differences, which are mentioned below, it is sufficient for our purposes to present a combined picture as in Table IV. The data are averaged separately for latencies of the first and second responses, T_1 and T_2. The rows of the matrices correspond to different values of t_1 and the columns to different values of t_2.

TABLE IV

LATENCY IN MILLISEC. AS A FUNCTION OF t_1 AND t_2 AVERAGED OVER SESSIONS AND SUBJECTS. PART II

T_1		*t_2 in millisec.*				
		50	100	160	230	310
t_1 in millisec. ..	50	283 (7)				303 (7)
	100		257 (5)		256 (4)	
	160			241 (3)		
	230		238 (6)		235 (6)	
	310	227 (6½)				231 (7)

T_2		*t_2 in millisec.*				
		50	100	160	230	310
t_1 in millisec. ..	50	303 (4)				188 (4)
	100		251 (3)		188 (2)	
	160			216 (2)		
	230		254 (3)		176 (2)	
	310	308 (2)				156 (0)

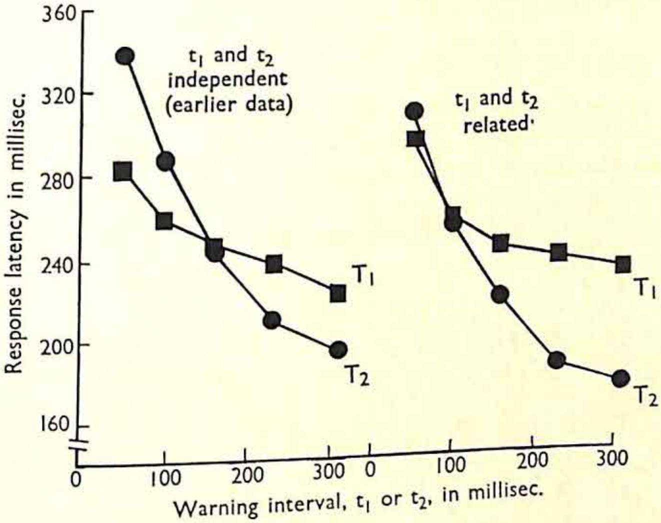
Note: the one-tail probabilities of the U scores shown above in brackets are given by the following p values:

U (or 12-U) ..	0	1	2	3	4	5	6
p	$\frac{1}{35}$	$\frac{2}{35}$	$\frac{4}{35}$	$\frac{7}{35}$	$\frac{11}{35}$	$\frac{15}{35}$	$\frac{20}{35}$

If the mean latencies in Table IV are compared with those in Table I then the following may be observed: for T_1 there is very good agreement between corresponding entries and, overall, latencies are 1·9 millisec. longer in the present data; for T_2, however, in every case of corresponding entries the latency in the present data is shorter, the average amount being 28 millisec. While tests have been carried out for differences in results between the two experiments none has rejected the null hypothesis either for T_1 or T_2. Since the comparison is between three subjects in one group and four in the other this is not surprising. In order to give some idea of the degree of overlap between experiments of the individual scores the Mann-Whitney U statistic was calculated for each cell in Table IV and the results are shown in brackets. Given the mean latencies of each of the seven subjects the U score is the number of pairwise interchanges of their results necessary to produce the ranking predicated on the assumption that latencies in the first experiment should be higher than those in the second. Thus a score $U = 0$ corresponds to complete agreement with this assumption while U = 12 corresponds to complete contradiction. The expected value of U on the null hypothesis is E (U) = 6 (Mood and Graybill, 1963, ch. 16). It is readily seen that whereas for T_1 the U scores in Table IV are distributed close to the expected value, for T_2 they are always below this value although seldom reaching significance.

The refractoriness effect may be depicted by averaging T_1 across t_2 and averaging T_2 across t_1. These results are shown in Figure 5, separately for the two experiments. In each graph T_1 is shown as a function of t_1 and T_2 as a function of t_2.

FIGURE 5



Latencies, T_1 and T_2, of the first and second responses as a function of the respective intervals, t_1 and t_2. Part II.

The major difference between the graphs is that the curves intersect at some intermediate delay in the earlier result and barely overlap in the present one. In fact, of the four subjects in the present experiment the results of only one showed any marked overlap, for another there was a small overlap at the shortest delay and for the other two there was no overlap.

In all subject-trials there were 45 occasions in which the response anticipated the stimulus and in all cases they occurred at the long delays of the relevant signal. Of these, 11 were first responses and 34 were second responses. The significance of these was discussed in Part I.

DISCUSSION

If the use of prediction is discounted and the result is taken as a replication of that in Part I, then it shows that transfer delays of the second response even at the small values of t_2 can be very small indeed or nonexistent.

If, however, there is an effect of redundancy and prediction is possible then what can we make of the issue as to whether this represents series or parallel processing of the input information? Suppose that the subject has a model with states A′, B′, C′ corresponding to states A, B, C of the stimulus source. Prediction of t_2 having observed t_1 requires that when the source is in state B_1 the model can attain state B_1' soon after S_1 appears. For the subject this is equivalent to having available the vector of probabilities of transition from B_1 to each of the C states.

It is theoretically possible for the model to be in B_1' at the same time as S_1 appears, providing that it can imitate the source in real time. The states in B (or in C) can be ordered as an ascending series of time intervals, and the correspondingly ordered states of B′ can be ticked off against elapsed time, so that, subject to accuracy of time estimation, the model is in B_1' if and when the source is in B_1. There is no question of overlapping processes because the process of prediction is built into the structure of the model.

It therefore seems consistent that prediction in the present task may be implemented in a serial processing system. This, however, assumes that the subject has access to a model that can imitate the source without appreciably lagging it in time. With other assumptions about access, for instance that the model does not participate in information processing but is retrieved piecemeal from memory as required, it would be more natural to consider overlapping processes. Thus with a small extension of the refractoriness task we find that the distinction between serial and parallel processing is too simple. It fails to make explicit some very important assumptions about the nature of the processing system.

REFERENCES

ANNETT, J. (1966). Payoff: a neglected factor in reaction time measurement. *Quart. J. exp. Psychol.*, **18**, 273–4.

BERTELSON, P. (1961). Serial redundancy and speed in a serial two-choice responding task. *Quart. J. exp. Psychol.*, **13**, 90–102.

BERTELSON, P. (1966). Central intermittency twenty years later. *Quart. J. exp. Psychol.*, **18**, 153–63.

BROADBENT, D. E. (1958). *Perception and Communication.* London: Pergamon.

DAVIS, R. (1956). The limits of the psychological refractory period. *Quart. J. exp. Psychol.*, **8**, 24–38.

EDWARDS, W. (1961). Costs and payoffs are instructions. *Psychol. Rev.*, **68**, 275–84.

GOTTSDANKER, R., BROADBENT, L., and VAN SANT, C. (1963). Reaction time to single and to first signals. *J. exp. Psychol.*, **66**, 163–7.

JERSILD, A. T. (1927). Mental set and shift. *Arch. Psychol., N.Y.*, **89**, 1–81.

KARLIN, L. (1966). Development of readiness to respond during short foreperiods. *J. exp. Psychol.*, **72**, 505–9.

MOOD, A. M., and GRAYBILL, F. A. (1963). *Introduction to the Theory of Statistics.* New York: McGraw-Hill.

MOWRER, O. H. (1940). Preparatory set (expectancy): some methods of measurement. *Psychol. Monogr.*, **52** (2, whole No. 233).

NICKERSON, R. S. (1965). Response time to the second of two successive signals as a function of absolute and relative duration of intersignal interval. *Percept. mot. Skills*, **21**, 3–10.

POULTON, E. C. (1950). Perceptual anticipation and reaction time. *Quart. J. exp. Psychol.*, **2**, 99–112.

SHAFFER, L. H. (1967). Transition effects in three-choice reaction with variable S–R mapping. *J. exp. Psychol.*, **73**, 101–8.

SMITH, M. C. (1967). Theories of the psychological refractory period. *Psychol. Bull.*, **67**, 202–13.

TELFORD, C. W. (1931). The refractory phase of voluntary and associative responses. *J. exp. Psychol.*, **14**, 1–35.

WELFORD, A. T. (1952). The "psychological refractory period" and the timing of high-speed performance—a review and a theory. *Brit. J. Psychol.*, **43**, 2–19.

WELFORD, A. T. (1967). Single-channel operation in the brain. *Acta Psychol.*, **27**, 5–22.

Manuscript received 8th August, 1968.

DISCRIMINABILITY AND CENTRAL INTERMITTENCY IN SAME-DIFFERENT JUDGEMENTS

BY

MICHAEL C. CORBALLIS*, WILLIAM LIEBERMAN†, and DALBIR BINDRA

From the Department of Psychology, McGill University

Four sets of paired visual stimuli (OO, XX, XO, or OX) were judged by 48 subjects to be either "same" or "different." Decision latencies of the same and different judgement were studied as a function of the inter-stimulus interval (ISI). In Experiments I and II, in which stimulus durations were 70 millisec., decision latencies showed marked increases when the ISI was reduced to 100 millisec., but in Experiments III and IV, in which the stimulus durations were only 40 millisec., comparable increases did not occur until the ISI was reduced to 50 millisec. These increases were more marked for "same" than for "different" judgements, although overall decision latencies were generally shorter for "same" judgements. The effects of varying ISIs and stimulus durations are interpreted in terms of masking; they fail to support ân hypothesis of central intermittency.

INTRODUCTION

Two recent studies of decision latencies for same-different judgements suggest that "same" and "different" latencies may have different characteristics. In particular, they may be differentially influenced by the discriminability of the stimuli to be judged. Bindra, Williams and Wise (1965), for example, found that subjects took *longer* to decide that two tones were of the same pitch than to decide they were of differing pitch; contrary to this, Nickerson (1965) reported *shorter* latencies for judging two visually presented letters the same than for judging them different. In both studies, the authors speculated that "same" latencies might become longer relative to "different" latencies the more difficult it is to discriminate the stimuli. This speculation receives some support from a comparison of the two studies. The tones presented by Bindra *et al.* (1965), when in fact different, were nonetheless closely similar, whereas the letters presented by Nickerson were easily discriminable. Moreover, in the former study, "same" latencies were even longer relative to "different" latencies when the discriminability of the tones was lowered by increasing the interval between them from 1 to 10 sec.

The present experiments investigate latencies for judging visually presented Xs and Os as same or different when the intervals between each pair are very brief, varying from 50 millisec. to 800 millisec. We might expect the stimuli to become more difficult to discriminate the shorter the inter-stimulus interval (ISI), because of masking, for example. As one consequence, decision latencies should increase. This follows from decision theory models of choice reaction time (e.g. Stone, 1960), which predict longer reaction times the lower the signal-to-noise ratio, and we may interpret reduced discriminability as a reduction in signal-to-noise ratio on some hypothetical dimension upon which same-different judgements are based. As a further consequence, if the speculations of Bindra *et al.* (1965) and Nickerson (1965) are correct, "same" latencies should increase more than "different" latencies.

There is another reason why decision latencies might increase when the ISI is very short. Davis (1964) found that latencies for judging visually presented Xs and

---

* Now at the University of Auckland, Auckland, New Zealand.

† Now at Washington University, St. Louis, Missouri, U.S.A.

Os same or different increased when the ISI was less than 300 millisec. He interpreted this as evidence that central analysis of stimuli is an intermittent process with a period of about 300 millisec., so that if a second stimulus arrives within 300 millisec. of a first it must wait in some short-term buffer storage until the central analysing mechanism is clear. This results in a delay to the subsequent decision. This hypothesis of central intermittency has also been suggested as an explanation of the so-called "psychological refractory period," the delay typically found in simple reaction time to a stimulus which occurs within about 300 millisec. of a previous stimulus (Davis, 1957, 1959; Welford, 1952).

Davis did not observe any differential effect of varying ISI upon "same" as compared to "different" latencies. However, in his experiments, the two stimuli of each pair were projected to different spatial locations, which should have eliminated any reduction in discriminability due to masking at short ISIs. In our experiments the stimuli were projected to the same spatial location.

We report four experiments. In Experiments I and II, subjects made "choice" judgements; that is, they made one of two possible responses depending upon whether they judged the stimuli same or different. In Experiments III and IV they made "recognition" responses; during some sessions they responded only if the stimuli were judged the same, withholding response if they were judged different, while during the other sessions they responded only if the stimuli were judged different. The ISIs, measured between onsets, ranged from 100 to 800 millisec. in Experiments I and II and from 50 to 200 millisec. in Experiments III and IV.

EXPERIMENT I

Subjects

Subjects were 16 volunteers from an introductory psychology class. Their ages ranged from 18 to 21 years.

Apparatus and Procedure

The stimuli were Xs and Os, projected by a multiple-stimulus projector (Grason-Stadler Model E-4580-3), with a 2 in. by 1½ in. screen. The durations of stimulus presentations and the intervals between stimuli were controlled by the output from a multichannel tape programming device; three channels were used to control presentations of the two stimuli and a fourth channel was used to activate two clocks.

The tape was pre-programmed to present four blocks of 37 trials. Each trial consisted of a 1-sec. warning tone presented through earphones, followed by one of the stimulus combinations: XX, OO, OX, or XO. Within each block, selection from among the four combinations on each trial was random except for the following restrictions: (1) after the first pair (which was omitted from all analyses) there were equal numbers of each alternative, and (2) there were equal numbers of same (S) pairs (i.e. XX or OO) following S pairs, different (D) pairs (i.e. OX or XO) following D pairs, S pairs following D pairs, and D pairs following S pairs. Each X and O was presented for 70 millisec. and presentation of the first of each pair began 1 sec. after offset of the warning tone. Two clocks were activated at the onset of the second stimulus of each pair.

The inter-stimulus interval (ISI)—the interval between onsets of the two stimuli of each pair—was varied between blocks but held constant within blocks. The four ISIs were 100, 200, 400, and 800 millisec. Order of presenting ISI blocks was balanced across subjects. To achieve this the subjects were divided into two groups of eight. Each group was assigned a different Latin Square, one consisting of the cyclic permutations of the intervals 100, 200, 800 and 400 millisec. and the other consisting of the cyclic permutations of 100, 400, 200 and 800 millisec. Thus two subjects received each of the different possible orders. Each subject was given a few practice trials at the first ISI he was to receive, before the experimental trials began.

Subjects viewed the projection screen at eye level from a distance on 3 ft. On an armrest to his preferred side were three microswitches arranged at the vertices of an isosceles triangle with the apex towards him. He was instructed to hold the index finger of his preferred hand on the nearest microswitch and, after presentation of a stimulus pair, to indicate whether the stimuli were the same or different as quickly as possible by lifting his finger and placing it on the appropriate one of the other two microswitches, a finger movement of about 1 in. For half the subjects the "same" microswitch was towards the side of his preferred hand, while for the remaining subjects the "different" microswitch was towards this side. One of the clocks which were activated when the second stimulus appeared stopped when the subject lifted his finger, recording what we shall call "premotor latency," and the other stopped when he pressed one of the decision microswitches, recording decision latency.

Results and discussion

Less than 2 per cent. of the total responses were errors. These errors are tabulated in Table I, along with those of Experiment II.

TABLE I

PER CENT. ERRORS MADE AT EACH INTER-STIMULUS INTERVAL (ISI) IN EXPERIMENTS I AND II

	ISI in millisec.			
	100	*200*	*400*	*800*
Experiment I	2·6	1·4	1·7	1·2
Experiment II	10·4	6·9	4·9	4·9

To analyse the latency trends, each subject's mean response-initiation and decision latencies for correct responses were calculated for each judgement ("same" or "different") at each ISI. The trends, shown in Figure 1, were assessed by analysis of variance, Latin Square design, carried out separately for each measure.

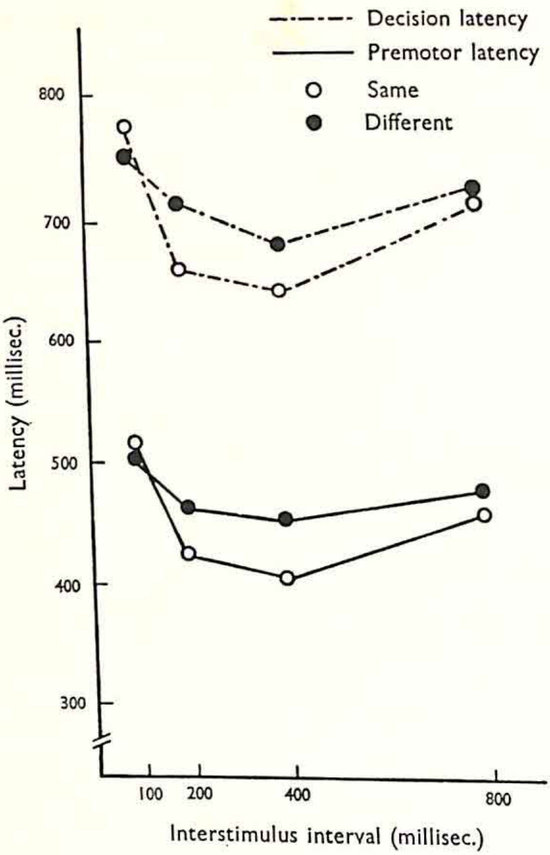
It is clear from Figure 1 that both premotor and decision latencies were longest when the ISI was 100 millisec. The main effect of ISIs was significant in both analyses; for premotor latency $F\ (3, 24) = 8{\cdot}05\ (p < 0{\cdot}001)$, and for decision latency $F\ (3, 24) = 8{\cdot}09\ (p < 0{\cdot}001)$. Individual comparisons among the means using the Newman-Keuls procedure (Winer, 1962, p. 80) indicated that both latencies were significantly ($p < 0{\cdot}05$) longer when the ISI was 100 millisec. than when it was 200, 400 or 800 millisec.

"Same" premotor and decision latencies were generally shorter than the corresponding "different" latencies, as in Nickerson's (1965) study. This difference was significant for premotor latency $F\ (1, 8) = 8{\cdot}08\ (p < 0{\cdot}05)$, but not for decision latency ($0{\cdot}05 < p < 0{\cdot}10$). For decision latency, however, there was a significant interaction between judgements and ISIs, $F\ (3, 24) = 3{\cdot}86\ (p < 0{\cdot}05)$. When the judgement effect was evaluated separately for each ISI by tests of "simple effects" (Winer, 1962, p. 174), it was found to be significant when the ISI was 200 and 400 millisec., but not when it was 100 or 800 millisec.; this was true for both premotor and decision latencies.

The only other significant effect was that of block positions on decision latency, $F\ (3, 24) = 3{\cdot}63\ (p < 0{\cdot}05)$. Decision latencies were longest for the first block; this probably resulted from a practice effect.

In general, these results are consistent with the view that the stimuli are least discriminable at the 100 millisec. ISI. Both premotor and decision latencies were the greatest when the ISI was 100 millisec.; further, "same" latencies increased more than did "different" latencies, as the results of Bindra, Williams and Wise (1965)

FIGURE 1



Premotor and decision latencies for "same" and "different" judgements at four inter-stimulus intervals obtained in Experiment I.

and Nickerson (1965) had led us to expect. There was also some tendency, more marked for "same" than for "different," for latencies to be longer at the 800 millisec. compared with 200 and 400 millisec. ISIs. This suggests that the optimal ISI for speed of performance may lie between 100 and 800 millisec.

Experiment II

Subjects

Sixteen volunteer subjects were drawn from the same source as in Experiment I.

Apparatus and procedure

The apparatus and stimuli were exactly the same as those used in Experiment I. Only the instructions differed.

The subjects were divided into two groups of eight. One group was instructed to make a response only if the stimuli were the same, and to withold response if they were different. The other group was instructed to respond only if the stimuli were different. The response apparatus was the same as in Experiment I. The subject was instructed to hold his finger on the nearest microswitch and, if a response was required, to press a specified one of the other two microswitches. Within each group half the subjects were instructed to press the microswitch towards the side of their preferred hand, while the other half were instructed to press the microswitch towards the other side. Thus only one of the microswitches was used for any given subject.

Order of presenting ISI blocks was balanced across subjects in a Latin Square which consisted of the cyclic permutations of the 100, 200, 800 and 400 millisec. ISIs. Each subject was given a few practice trials at the first ISI he was to receive.

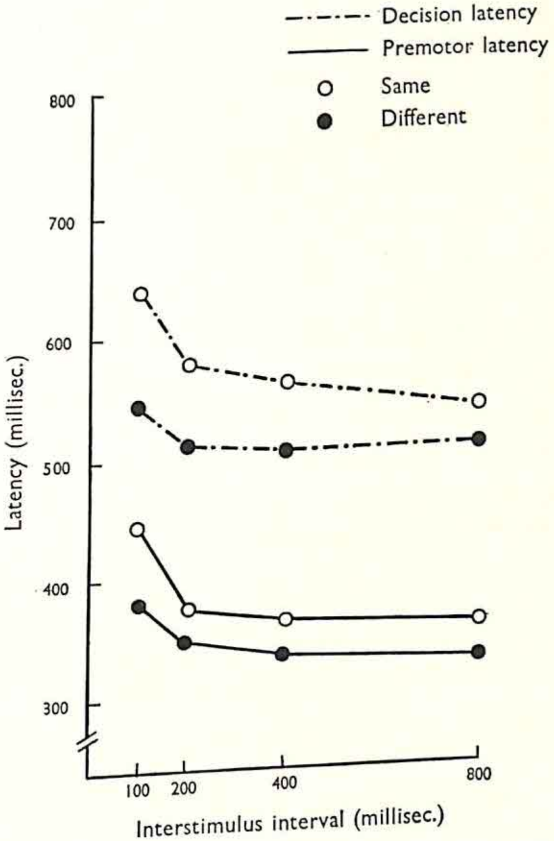
Results and discussion

There were again very few errors. They are tabulated in Table I.

Mean premotor and decision latencies for correct responses were calculated for each subject at each ISI and subjected to analyses of variance, Latin Square design. The only significant effect was that due to ISIs; for premotor latency F (3, 24) $=$ 19·05 ($p <$ 0·001), and for decision latency F (3, 24) $=$ 9·24 ($p <$ 0·001). Individual comparisons among the means using the Newman-Keuls procedure indicated that both latencies were significantly ($p <$ 0·05) longer when the ISI was 100 millisec. than when it was 200, 400, or 800 millisec. This is illustrated in Figure 2.

Figure 2 also shows that the increase in premotor and decision latencies at the 100 millisec. ISI was greater for "same" than for "different" judgements. Although this interaction was not significant in either analysis, it was similar to that observed in Experiment I. However, the fact that premotor and decision latencies were both *longer* overall for "same" than for "different" judgements (although not significantly so) is in direct contrast to the results of Experiment I. This comparison was a "between subjects" comparison, so that this result may have been due to selection bias; it seems unlikely that there is any fundamental difference between choice and recognition responses which would produce such a reversal. Nevertheless the point required further investigation. In Experiments III and IV, therefore, we repeated

FIGURE 2



Premotor and decision latencies for "same" and "different" judgements at four inter-stimulus intervals obtained in Experiment II.

the main features of Experiments I and II, respectively, but the experimental design of Experiment IV (unlike that of Experiment II) was such that "same" versus "different" was a "within subjects" comparison. In Experiments III and IV, too, the stimulus durations and ISIs were made shorter than in Experiments I and II.

EXPERIMENT III

Subjects

The subjects were eight students employed for summer research. Their ages ranged from 16 to 20 years.

Apparatus and procedure

The stimuli, projector and screen were the same as in Experiments I and II. Stimulus presentations and durations between stimulus onsets were controlled by five timers.

There were four blocks of trials. Each subject was given 10 practice trials before the first block and five practice trials before each of the remaining three. Each block consisted of 44 experimental trials. A 1-sec. warning tone presented through earphones preceded each stimulus pair, as in the first two experiments. On each trial, the experimenter selected the stimulus pair from among the alternatives XX, OO, OX and XO according to a pre-arranged schedule, random except for the restrictions described for the stimulus presentations in Experiment I. Each stimulus was presented for 40 millisec. Note that this is shorter than in the previous experiments, in which each stimulus was presented for 70 millisec.

The ISIs were 50, 100, 150 and 200 millisec. The ISI was constant within each block but varied between blocks according to a Latin Square design, the cyclic permutations of 50, 150, 200 and 100 millisec. Two subjects were assigned to each order.

The instructions, response apparatus, and method of recording premotor and decision latencies were the same as in Experiment I. Subjects made choice responses, as in Experiment I.

Results and discussion

Errors are tabulated in Table II. Mean premotor and decision latencies were calculated as in Experiment I and subjected to analyses of variance, Latin Square design; the trends are illustrated in Figure 3.

TABLE II

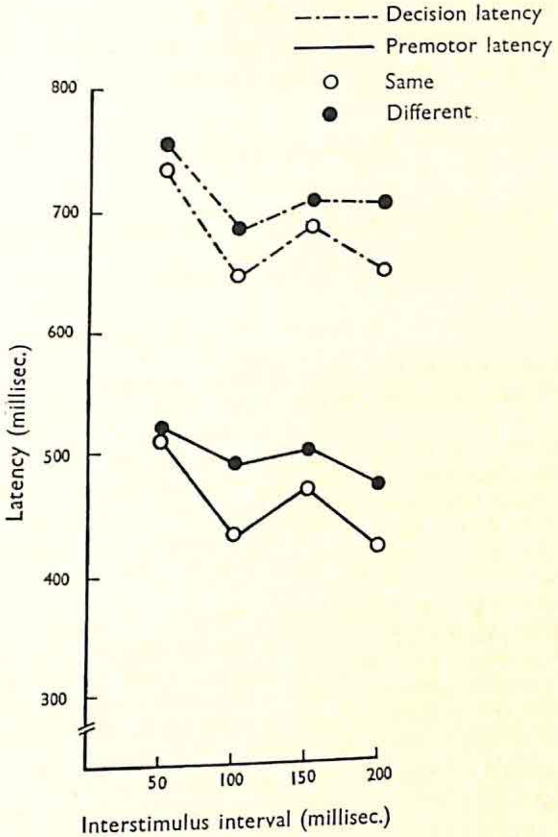
PER CENT. ERRORS MADE AT EACH INTER-STIMULUS INTERVAL (ISI) IN EXPERIMENTS III AND IV

	ISI in millisec.			
	50	*100*	*150*	*200*
Experiment III	6·0	2·0	1·5	1·2
Experiment IV	13·7	11·9	6·8	8·0

In both analyses the main effect due to ISIs was significant; for premotor latency F (3, 12) = 6·19 (p <0·01), and for decision latency F (3, 12) = 17·48 (p <0·001). Figure 3 shows that the longest latencies occurred at the 50 millisec. ISI. Contrary to the previous results, there was no evidence of an increase in latencies when the ISI was 100 millisec.

Figure 3 shows that the latencies were somewhat longer when the ISI was 150 millisec. than when it was 100 or 200 millisec. This effect was almost entirely due to the two subjects who received the 150 millisec. ISI first; they in fact had longer

FIGURE 3



Premotor and decision latencies for "same" and "different" judgements at four inter-stimulus intervals obtained in Experiment III.

latencies at the 150 millisec. ISI than at the 50 millisec. ISI. The first block of trials, as compared to the remaining three blocks, yielded longer latencies. The main effect due to block position was significant for decision latencies F (3, 12) = 26·73 (p <0·001). This may be attributed partly to practice, but this is not the whole explanation, since the within subjects component of the interaction between block positions and ISIs was also significant, F (6, 12) = 14·00 (p <0·001). We do not know why the latencies should have been particularly long at the 150 millisec. ISI when it was presented first, but if we discount this effect the picture which emerges is that the latencies were virtually the same when the ISIs were 100, 150 and 200 millisec., and were considerably longer when the ISI was 50 millisec.

"Same" latencies were shorter than "different" latencies as in Experiment I and in Nickerson's (1965) study. This difference was significant in both analyses; for premotor latency F (1, 4) = 58·8 (p <0·01), and for decision latency F (1, 4) = 21·59 (p <0·01). The interaction between judgements and ISIs was also significant in both analyses; for premotor latency F (3, 12) = 6·38 (p <0·01), and for decision latency F (3, 12) = 4·12 (p <0·05). Tests of simple effects indicated that the judgement effect was not significant when the ISI was 50 millisec., but was significant at the other three ISIs, in both analyses.

To summarize, then, the results can be interpreted as indicating that the stimuli were most difficult to discriminate when the ISI was 50 millisec. Both latencies increased at this ISI, and this increase was more marked for "same" than for "different" judgements. The principal difference between these results and those of Experiment I was that these effects did not occur when the ISI was 100 millisec., as they did in Experiment I. The reason for this discrepancy is probably that the stimulus durations were shorter in this experiment than in Experiment I. We shall return to this point in General Discussion.

EXPERIMENT IV

Subjects

The subjects were eight students employed for summer research. Their ages ranged from 15 to 19 years.

Apparatus and procedure

The apparatus was the same as that used in Experiment III. Stimulus presentations were also the same, except that each subject received the four blocks of trials twice, on consecutive days.

The instructions to the subjects were the same as in Experiment II; that is, they were instructed either to make a response if the stimuli were the same and to withhold response if they were different ("same only"), or to make a response only if the stimuli were different ("different only"). The subjects were divided into two groups of four; one group received "same only" instructions on the first day of testing and "different only" on the second day, while the other group received "different only" on the first day and "same only" on the second. On each day, each group received the four ISIs. The ISIs were balanced across subjects in a Latin Square design, which consisted of the cyclic permutations of 50, 150, 200 and 100 millisec.

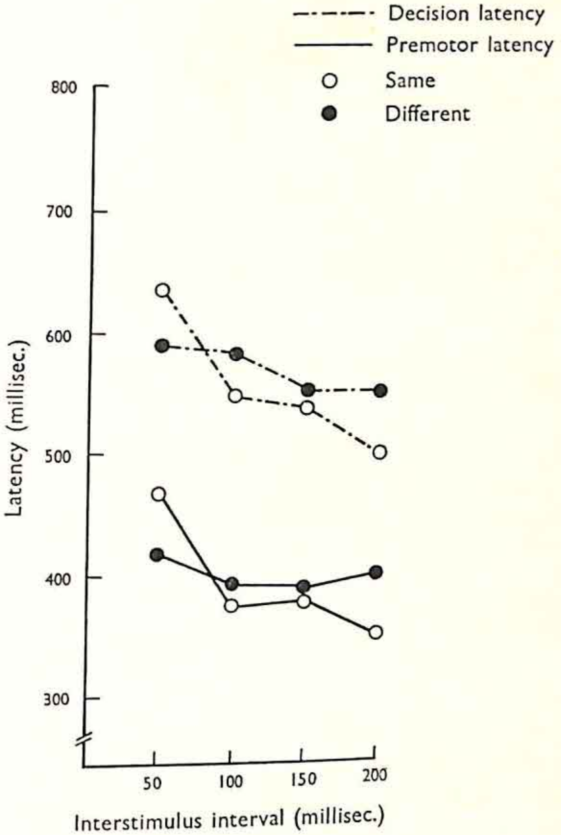
Results and discussion

The errors are tabulated in Table II. Mean premotor and decision latencies were calculated for each subject at each ISI for each day. Analyses of variance of these data posed a problem. The experimental design consists of *two* Latin Squares, one to balance the order of receiving the different ISIs, the other to balance the order of receiving different instructions ("same only" and "different only"). If the variance is fully analysed there are no error terms for testing significance, since each subject received a unique sequence of treatments. In carrying out analyses of variance, therefore, we ignored possible order effects associated with presentation of ISIs, effectively treating ISIs as a simple repeated measures variable. As a result, our tests of significance should be too conservative, if anything, since the effects of different orders or block positions might be expected to inflate the error terms but not the main effects of our analysis. We retained the second of the two Latin Squares in our analysis of variance, however, permitting a comparison of scores between the two testing days.

The main effect due to ISIs was significant in both analyses; for premotor latency $F(3, 18) = 12{\cdot}19$ ($p < 0{\cdot}001$), and for decision latency $F(3, 18) = 10{\cdot}90$ ($p < 0{\cdot}001$). It is clear from Figure 4 that latencies were the longest at the 50 millisec. ISI. Comparisons between means using the Newman-Keuls procedure indicated that mean latencies at the 50 millisec. ISI were significantly ($p < 0{\cdot}01$) longer than those at the other three ISIs. In addition, mean decision latency at the 100 millisec. ISI differed significantly ($p < 0{\cdot}01$) from those at the 150 and 200 millisec. ISIs.

The judgement effect was not significant in either analysis. However the interaction between judgements and ISIs was significant in both; for premotor latency $F(3, 18) = 23{\cdot}92$ ($p < 0{\cdot}001$), and for decision latency $F(3, 18) = 9{\cdot}03$ ($p < 0{\cdot}001$). The nature of this interaction is illustrated in Figure 4. It is clear that "same" judgements became increasingly longer relative to "different" judgements as ISIs decreased.

FIGURE 4



Premotor and decision latencies for "same" and "different" judgements at four inter-stimulus intervals obtained in Experiment IV.

Decision latencies were significantly shorter on the second day than on the first F (1, 6) = 8·42 (p <0·05). This can be interpreted as a practice effect.

To summarize, the results are once again consistent with the interpretation that discriminability decreased as the ISIs decreased. As in the previous experiment, the striking contrast was between the 50 millisec. ISI on the one hand, and the 100, 150 and 200 millisec. ISIs on the other. Though there was some evidence that decision latencies, but not premotor latencies, were longer at the 100 millisec. ISI compared with those at the 200 millisec. ISI, this comparison was not so marked as it was in Experiment II. We shall suggest that a critical difference between this experiment and Experiment II was that stimulus durations were shorter in this experiment.

GENERAL DISCUSSION

First, it is clear that both premotor and decision latencies showed virtually the same trends, indicating that subjects made the decisions before initiating the finger movement. If there is a discrepancy between the two, it is that practice effects were rather more marked for decision latencies than for premotor latencies, suggesting that practice influenced the motor components rather than the premotor components of decision time. But the other decision latency trends were apparent in the premotor latency as well, and so we infer that these trends reflect characteristics of premotor components common to both measures. To avoid repetition, therefore, we shall use the term "decision latency" to refer to these common, premotor components.

Our results confirm that the decisions "same" and "different" have different latency characteristics. In Experiments I and III, where subjects made "choice" judgements, "same" decisions were generally faster than were "different" decisions. This is consistent with Nickerson's (1965) result but not with that of Bindra *et al.* (1965), who reported the opposite result. If, however, as these authors suggest, discriminability is the important variable determining relative speed of "same" and "different" decisions, then our result is as expected. The stimuli we used were visual and easy to discriminate, like Nickerson's, but unlike those used by Bindra *et al.* (1965), which were auditory and difficult to discriminate. However, the results of Experiments II and IV, in which subjects made "recognition" responses, are more equivocal. In Experiment II "same" decisions were the slower, though not significantly so. This could have been due to sampling bias. In Experiment IV "same" decisions were the faster at all but the 50 millisec. ISI, but again the difference was not significant.

Of more interest than the overall differences between "same" and "different" latencies were the interactions between judgements and ISIs, significant in all but Experiment II. In each experiment the increases in latency at the shortest ISI were more marked for "same" than for "different" judgements. If it is assumed that discriminability was lowered at the shortest ISIs, this can be taken as further confirmation that reducing discriminability increases "same" latency more than "different" latency. In terms of decision theory models of choice reaction time (e.g. Stone, 1960), it appears that the criterion for making "same" decisions becomes relatively more stringent than the criterion for making "different" decisions as discriminability decreases. This interpretation has been elaborated by Bindra *et al.* (1965).

It is likely that reduced discriminability at the shortest ISIs was due to masking. One reason for supposing this to be so is that the increases in latency were a function of stimulus duration as well as of ISI. Fraisse (1966) has shown that, for pairs of letters presented in rapid succession, masking effects become more pronounced as stimulus durations increase. Similarly, in our experiments increases in decision latency were more marked the longer the stimulus duration. In Experiments I and II, in which the stimulus durations were 70 millisec., decision latencies increased when the ISI was reduced to 100 millisec., but in Experiments III and IV, in which the stimulus durations were only 40 millisec., comparable increases in decision latency were not observed until the ISI was reduced to 50 millisec. Note also that errors were generally more frequent at the shortest ISIs (see Table I), though they were too few for comparisons to be reliable.

Our results differ in certain important respects from those of Davis (1964), although they are superficially similar in that they show increased latencies for short ISIs. For example, Davis observed a reversal effect at the very shortest ISI under a "regular interval" condition in which ISIs remained constant within blocks of trials. Under this condition reaction times were longest at the second shortest ISI (160 millisec.) and *decreased* again at the shortest ISI (35 millisec.). This reversal effect did not occur under a "random interval" condition in which ISIs varied randomly from trial to trial; under this condition decision latencies were longest at the 35 millisec. ISI. The presentation conditions in our experiments corresponded to Davis's regular interval condition, yet we did not observe any reversal effects, even though the temporal parameters of Experiments III and IV were similar to those of Davis's experiment. The reversal effect is evidence that masking was not a factor in Davis's experiment, and that discriminability was not reduced at short ISIs. Indeed, since the stimuli of each pair were presented to different spatial locations, one would not expect masking to have occurred. Davis explained the reversal effect in terms of central intermittency by arguing that when the ISI is extremely short, and when (as in the regular interval condition) the subject knows in advance that it will be short, he can treat each pair of stimuli as a single perceptual unit and so need not analyse them serially.

Another difference between our results and Davis's is that much shorter ISIs were required to produce latency increases in our experiments than in his. Davis found increased latencies when the ISI was 160 millisec. and each stimulus was presented for 35 millisec. This contrasts with the results of Experiment III of our study, in which latency increases did not occur until the ISI was reduced to 50 millisec. In general, there was no suggestion at all from our results that latency increases were related to central intermittency with a period of 300 millisec. Davis's results, by contrast, cannot be explained in terms of masking but are wholly consistent with the hypothesis of central intermittency. It is possible that intermittency occurs only when the stimuli are presented to different spatial locations, as in this study by Davis, or when they are presented to different sense modalities (Davis, 1957). These experimental situations may involve a switch of attention, in some meaning of the phrase, between stimuli. On the other hand, if the stimuli are presented to the same spatial location, as in our experiments, they may be automatically "grouped" into a single perceptual unit for all ISIs less than 300 millisec., and not just when ISI is very short. Whether this is so in other situations, for example, when same-different judgements are not involved, or when ISIs vary randomly from trial to trial so that the subject does not know in advance what ISI he will receive, remains to be investigated.

This work was supported by a contract (Nonr-4896(00)) between the U.S. Office of Naval Research and McGill University, and by a grant from the Defence Research Board of Canada.

References

Bindra, D., Williams, J. A., and Wise, J. S. (1965). Judgments of sameness and difference: Experiments on decision time. *Science*, **150**, 1625–7.

Davis, R. (1957). The human operator as a single channel information system. *Quart. J. exp. Psychol.*, **9**, 119–29.

Davis, R. (1959). The role of "attention" in the psychological refractory period. *Quart. J. exp. Psychol.*, **11**, 211–20.

Davis, R. (1964). The combination of information from different sources. *Quart. J. exp. Psychol.*, **16**, 332–9.

Fraisse, P. (1966). Visual perceptive simultaneity and masking of letters successively presented. *Percept. Psychophys.*, **1**, 285–7.

Nickerson, R. S. (1965). Response times for "same"-"different" judgments. *Percept. mot. Skills*, **20**, 15–8.

Stone, M. (1960). Models for choice reaction-time. *Psychometrika*, **25**, 251–60.

Welford, A. T. (1952). The "psychological refractory period" and the timing of high speed performance: A review and a theory. *Brit. J. Psychol.*, **43**, 2–19.

Winer, B. J. (1962). *Statistical Principles in Experimental Design.* New York: McGraw-Hill.

Manuscript received 27th June, 1967.

SELECTION IN VISUAL IMMEDIATE MEMORY

BY

J. M. VON WRIGHT

From the Department of Psychology, University of Turku, Finland

The sampling (partial report) technique of Sperling (1960) was used to study the efficiency of selection of letters from visual immediate memory (sensory storage). Selection by location, chromatic colour, achromatic colour, and size was fairly efficient whereas selection by orientation was difficult. Some parallels between selection in visual immediate memory and selective listening are briefly discussed.

INTRODUCTION

(1) In the models proposed during recent years to account for the experimental data on short-term memory and selective attention, a distinction is usually made between two kinds or stages of immediate memory: a sensory or pre-perceptual storage which decays rapidly with time, and a post-perceptual storage of substantially longer duration (e.g. Broadbent, 1963; Melton, 1963; Mowbray, 1964; Murdock, 1966). The former will here be called VIS (visual information storage) when visual stimuli are used, the latter being called STM (short-term memory). Information available in VIS decays rapidly unless it is processed—read out, responded to—the maximal storage time with a bright post-exposure field being 200–300 millisec. (Averbach and Sperling, 1961; Sperling, 1963). Uncategorized auditory signals appear to be available somewhat longer, approx. 1–2 sec. (Treisman, 1964*a*).

Sperling (1960) has demonstrated that significantly more information may be available in VIS than can be "read out" from it and reported. His technique was, briefly, as follows. The subject was shown tachistoscopically an array of items and asked to recall a part of the array, the signal specifying the part to be reported being given after the termination of the visual presentation. This method samples the information that the subject has available when the selection signal has been processed. The total amount of information available can be estimated from the sample. This estimate, called *partial report*, may be compared with the number of items the subject can report in a traditional immediate-memory test (*whole report*). The difference between the partial and whole reports provides an estimate of the amount of information that was available in excess of that which could be reported; it also provides a crude measure of the efficiency of the selection process.

For the present purposes, Sperling's procedure can be described in general terms as follows. The subject is instructed to perform two successive tasks: (*a*) to select (from VIS), according to a specified criterion, a part of the stimulus array, and (*b*) to report certain characteristics of the selected material. In all but one of the experiments the subject was instructed to select the row or column indicated by a signal, and to report the letters in that row (column). Thus *location* was used as *selection criterion*, and *letters reproduced in correct position* as *response criterion*. In these as well as in subsequent experiments selection by location has proved efficient, so long as the selection signal is given within a fraction of a second after the termination of stimulus presentation. In favourable conditions partial reports three times greater than the corresponding whole reports have been obtained (Sperling, 1963).

In one experiment, however, Sperling (1960, Exp. 6) used a different selection criterion: the stimulus array consisted of letters and numbers randomly intermixed, and the subject's task was to report either the letters or the numbers. In this case selection was difficult: the subjects' partial reports were only negligibly more accurate than their whole reports of the same material.

(2) It has been shown in experiments on selective listening that subjects are able to respond efficiently to one of two simultaneous auditory messages if these differ in general physical features such as spatial position, whether real or apparent, frequency range, and intensity (see review in Treisman, 1964*b*). Selection on the basis of the verbal content of the messages, on the other hand, is difficult so long as the general physical characteristics of the messages are similar, even if the two messages are, e.g. in different languages known to the subject. These results have been taken to support the type of model for selective attention which has been proposed by Broadbent (1958). This model postulates a "selective filter" which passes or rejects messages primarily on the basis of their general physical characteristics, prior to the "full analysis" of any message.

The Sperling type of experiment differs in at least two important respects from most studies on selective listening: in the former the subject is given a "single burst of information," the selection signal usually being given after the stimulus presentation, whereas in the latter he has to respond to a continuous information input and is instructed in advance which message he has to select. In spite of these differences, however, it would seem to be of some interest to study the extent to which the characteristics of selective listening (including the fate of the rejected messages) have counterparts in the selection of material from VIS.

(3) The present study is concerned with the efficiency of selection in VIS when different selection criteria are used. In the experiment reported below, location, chromatic and achromatic colour, size, and orientation are used as such criteria. The experiment differs in one crucial respect from the long series of studies in which subjects have been asked to attend to and report selectively various characteristics of a stimulus array (e.g. Külpe, 1904; Chapman, 1932; Harris and Haber, 1963). In the latter the response criterion was systematically varied, whereas in the present experiment it is kept constant. The subject's task is in each case to report letters in correct position; only the criterion according to which the letters are selected is varied.

GENERAL METHOD

The experimental situation was essentially the same as in Sperling's (1960) study. The apparatus was a two-field mirror tachistoscope with a mechanical timer. Viewing was binocular at a distance of 50 cm. The fixation field was dimly illuminated with a fixation point in the middle.

The subject was instructed to look at the fixation point until it was clearly in focus; he then pressed a button which initiated the presentation. The exposure time was 100 millisec. Five millisec. after the termination of exposure a selection signal was delivered to the subject through earphones. The signal consisted of a short pip of either 1500 cps or 300 cps.

Stimulus material. The stimuli were lettered 20 × 16 cm. cards with either two rows of four letters each (4/4), or three rows of three letters each (3/3/3). The letters were consonant capitals drawn in indian ink. They were 10 mm. high, the distance between the midpoints of any two adjacent letters being 15 mm. No subject was required to report the same part of any stimulus card more than three times, discounting the stimuli used for training. In all, over 300 different cards were used.

Subjects. Fourteen undergraduates served as subjects and were paid for this by the hour. Each subject participated in six 1-hr. sessions. Four subjects were replaced by new ones, since they found it difficult to adapt to the experimental situation, their whole reports being consistently less than three letters.

Procedure. Each subject participated in four different experimental conditions. Each condition included (1) series of presentations in which the subject was instructed to report as many letters as possible (whole report), and (2) series of presentations in which only letters specified by the selection signal were to be reported (partial report). In the latter case the subject was told that a tone would be sounded approximately simultaneously with the exposure. If it was a high tone he was to report all the letters having a certain characteristic (e.g. all green letters, or all small letters). If it was a low tone, he had to report all the letters having another characteristic.

Series of 10–25 trials were grouped together without change in conditions. Whenever conditions were changed, a number of preliminary trials were given to make sure that the subject interpreted the choice signals correctly. No systematic changes in performance in any conditions were found after the second session. In the final analysis, only the results from the last three sessions were included. During these sessions the rate of stimulus presentation, determined by the subject himself, was on the average three stimuli per minute.

The responses were recorded on specially prepared response sheets. These were placed on the table immediately below the tachistoscope. The subject was instructed to write down the letters immediately on hearing the choice signal, the room illumination being adequate for writing.

Scoring. Every report was scored both for the total number of letters which agreed with those in the stimulus array and for the number of letters reported in their correct position. These two scores were closely parallel for all subjects but one. The second of these scores (letters in correct position) are tabulated in Table I. For the *whole report* condition these scores are given as such. For the *partial report* condition a procedure corresponding to that used by Sperling (1960, p. 7) was employed in order to obtain an estimate of the letters available to the subject: the proportion of correctly reported letters out of all letters in the category to be selected was multiplied by the total number of letters on the card.

EXPERIMENTAL CONDITIONS

Five different conditions were used.

I. Selection by location. The subjects were instructed to report the letters in the top row or bottom row of 4/4 cards, depending on the pitch of the selection signal. This condition is a replication of one of Sperling's (1960) experiments, and was included as a control condition.

II. Selection by colour. Two colours were used for the letters on each stimulus card. In the 4/4 cards half of the letters were of one and the other half of another colour. In the 3/3/3 cards four letters were of one and five of another colour. The location of the colours was randomly varied. Red, blue, and yellow were used as colours. Each subject was given an equal number of stimulus cards with each of the three two-colour combinations. A detailed analysis of the results showed that there were no significant or clear-cut differences between the different colour pairs; consequently the scores were combined over colours.

III. Selection by achromatic colour. The letters on the stimulus cards were drawn in two different shades of grey, approximately corresponding to colour proofs 12 (reflectance 23 per cent.) and 23 (reflectance 8 per cent.) of the grey-scale in Hesselgren's Colour Atlas (1952).

IV. Selection by size. The letters on the stimulus cards were drawn in two different sizes, 10 mm. high and 6 mm. high, respectively. The difference between the midpoints of any two adjacent letters was 15 mm., independent of the size of the letters.

V. Selection by orientation. Two types of stimulus cards were used. *A:* half of the letters were drawn in an upright position, the other half being turned 45 degrees to the right (−45°). *B:* half of the letters were drawn turned 45 degrees to the right (−45°), the other half being drawn turned 45 degrees to the left (+45°). The subjects were told to report (reproduce) the letters in any orientation they found convenient. In fact they were almost invariably reproduced in the usual upright position.

RESULTS

The results are shown in Table I . In *Condition I*, the partial reports are significantly higher than the whole reports for each subject ($t > 6{\cdot}0$ in each case). Both the whole and the partial reports are on the average lower than those obtained by Sperling (1960, Table 3). This might be due to uncontrolled differences between the experimental situations. However, the intra-individual differences between the two conditions are similar in Sperling's and in the present study, indicating that selection by location is highly efficient.

In *Condition II* the difference between partial and whole reports is highly significant ($p < 0{\cdot}001$) for 10 out of 12 subjects; for several subjects the distributions of the two types of report do not even overlap. The results indicate that letters can be selected with fair efficiency from VIS when (chromatic) colour is used as selection criterion.

In *Condition III* the difference between partial and whole reports is highly significant for two subjects out of four, and barely significant for one. In *Condition IV* this difference is significant for all subjects, although barely so in one case. In the former condition only two shades of grey, and in the latter condition only two letter sizes were used. However, the results suggest that selection by "greyness" and by letter size can occur with some efficiency, although it appeared to be more difficult than selection by chromatic colour.

In *Condition V* none of the intra-individual differences between the two types of report even approached significance. It may be noted, however, that with the *B*-cards the whole reports were somewhat greater than the partial reports for all subjects. With the *A*-cards all subjects reproduced more "turned" (−45°) than upright letters in the partial report condition, but none of the differences were significant. With both types of cards, approximately the same number of letters in each orientation was reported by all subjects in the *whole* report condition, the subjects almost invariably adopting a selection-by-location strategy (see below). Thus the lack of differences between the two types of report can hardly have been due to any differential difficulties in perceiving or reporting letters in different orientations.

The lack of clear-cut or consistent differences between the partial and whole reports indicates that selection of letters from VIS on the basis of their orientation is difficult.

An inspection of Table I shows that the inter-individual differences were fairly consistent over conditions: subjects who performed efficiently in one condition tended to do so in other conditions, too. Two subjects (b and d) had comparatively low partial report scores in all conditions. The former remembered on the average the same number of correct letters as the other subjects but made a great number of location errors. The latter subject changed her observation strategy repeatedly.

DISCUSSION

(1) To the extent that significant differences between a subject's partial and whole reports of the same stimulus array can be taken as evidence of efficient selection, the present results show that selection of letters from an array can occur efficiently with location, chromatic colour, achromatic colour, and size as selection criteria. Selection by orientation, on the other hand, appears to be difficult with the present type of stimuli.

No detailed conclusions can be drawn from the data obtained in Conditions III and IV since only two degrees of grey and two letter sizes, respectively, were used;

TABLE I

AVERAGE WHOLE REPORTS AND PARTIAL REPORTS

Selection criterion	*Subject*	*Stimulus card*	*n*	*Partial report*	*Whole report*	*Diff.*	*p*
I. Location	f	4/4	20	5·0	3·1	1·9	0·001
	g	4/4	20	6·8	4·0	2·8	0·001
	h	4/4	30	6·9	4·2	2·7	0·001
	i	4/4	30	6·9	4·4	2·5	0·001
	j	4/4	30	6·5	4·0	2·5	0·001
	k	4/4	30	6·9	4·1	2·8	0·001
	Mean			6·5	4·0	2·5	
II. Colour	a	4/4	30	6·2	3·0	3·2	0·001
	b	4/4	30	4·0	3·4	0·6	—
	c	4/4	30	6·3	3·2	3·1	0·001
	d	4/4	30	4·6	3·8	0·8	0·10
	e	4/4	30	7·0	3·7	3·3	0·001
	f	3/3/3	60	5·8	3·0	2·8	0·001
	g	3/3/3	60	6·4	3·8	2·6	0·001
	h	3/3/3	60	6·6	4·1	2·5	0·001
	i	3/3/3	60	5·2	3·6	1·6	0·001
	j	3/3/3	60	6·1	3·7	2·4	0·001
	k	3/3/3	60	5·2	3·6	1·6	0·001
	l	3/3/3	60	5·4	3·5	1·9	0·001
	Mean			5·7	3·5	2·2	
III. Achromatic colour	a	4/4	20	4·9	3·3	1·6	0·001
	b	4/4	20	5·0	4·3	0·7	—
	c	4/4	20	5·3	3·4	1·9	0·001
	d	4/4	20	4·4	3·3	1·1	0·05
	Mean			4·9	3·6	1·3	
IV. Size	a	4/4	20	5·2	3·1	2·1	0·001
	b	4/4	20	4·5	3·6	0·9	0·05
	c	4/4	20	5·2	3·6	1·6	0·001
	d	4/4	20	4·7	3·4	1·3	0·01
	Mean			4·9	3·4	1·5	
V. *A*. Orientation 0° vs. −45°	a	4/4	20	4·4	3·8	0·6	—
	b	4/4	20	4·1	3·7	0·4	—
	c	4/4	20	4·6	4·2	0·4	—
	d	4/4	20	3·1	3·7	−0·6	—
	Mean			4·0	3·8	0·2	
V. *B*. Orientation −45° vs. +45°	a	4/4	10	3·2	3·8	−0·6	—
	b	4/4	10	3·8	4·2	−0·4	—
	c	4/4	10	3·7	4·2	−0·5	—
	d	4/4	10	2·6	3·2	−0·6	—
	Mean			3·3	3·8	−0·5	

n = Number of observations from which the individual means were calculated.

the results only show that fairly efficient selection *can* occur. Some pilot trials with different size-pairs suggested, as one might expect, that the ease of selection by size is a function of the (subjective) difference between the sizes involved, but there is no precise information on this relationship.

The intra-individual variance of the *partial* reports was on the whole inversely related to the average accuracy of these reports in Conditions I–IV: the variance was least in Condition I and greatest in Condition III. In the last mentioned case almost 5 per cent. of the responses were complete failures, i.e. contained no correct letters, whereas such failures never occurred in Condition I, and occurred only in 2/12 subjects in Condition II. The variance of the whole reports was small throughout and no complete failures occurred. In fact, the intra-individual variance appeared to be dependent on the observation strategies adopted by the subjects.

The observation strategies of the subjects are obviously important for the outcome of experiments of the present type. Anticipation of a particular selection signal, i.e. a guessing strategy, was more frequent in the first session than later, the subjects apparently becoming convinced that it was impossible to outguess the experimenter. However, when selection was found difficult (for instance in Condition III), subjects occasionally reverted to guessing. This greatly increased the variance (and led to occasional complete failures), and tended to make for a decrease in the mean of the correct responses in a sequence of trials. Sperling (1960) reports a corresponding increase in the use of the guessing strategy when the instruction (selection signal) was delayed.

Several subjects described their usual strategy during late sessions as one of "just waiting for the signal." An incidental observation suggests that this description may have been very apt indeed: for two subjects the selection signal was once omitted during the last (sixth) session in a partial report trial-series. In neither case could the subject report anything at all of the stimulus array except that it had been visible.

(2) The differences between the partial and whole reports were greater when selection was by location than when other criteria were used. At least two factors contributed to this:

(*a*) When a subject has to select a row of letters from an array with several rows, he usually reads it out from left to right, and writes it down on the response sheet in the same order. Even if the letters are not read out in order, there are comparatively few possible locations for each remembered letter in the row. With selection criteria other than location the letters to be picked out were randomly distributed around the array, and the uncertainty as to the location of each remembered letter was much greater; locating the intended position on the response sheet also took more time. In fact several subjects in the present study complained about the difficulty of remembering the location of the letters. If we tabulate the number of letters correctly reproduced *without regard to position*, the mean partial report scores of all subjects in Conditions I–IV are significantly higher than the corresponding whole report scores.

(*b*) Selection by location was judged to be particularly easy, and was spontaneously adopted when there were no instructions to the contrary. In the whole report condition the subjects almost invariably reported either a row or a block of letters, and when "stray" letters outside this row or block were included in the report they were incorrect as often as not.

(3) The experiments by Sperling (1960) and by Averbach and Sperling (1961) in which the delay before the selection signal was systematically varied show that the contents of VIS decay rapidly. Thus we may assume that the faster and simpler the process involved in "selecting items according to a specified criterion," the greater is the number of items that can be selected for further processing within the limited time available. That is to say, the difference between the partial and the whole

reports is a function of how rapidly the initial selection occurs, and may give some information about the time it takes to decide whether or not a given item belongs to a specified category.

Sperling found selection to be difficult when letters or numbers had to be reported selectively from an array containing both (see above). This could be accounted for by the fact that it is difficult to distinguish between (written) capital letters and numbers on the basis of their general physical characteristics; thus each item in the array might have to be identified—analysed in full—before categorizing can occur. It is possible that the same is true for the items in the present Condition V, i.e. that the letters had to be identified before their orientation could be determined, since letter-shapes contain a variety of differently oriented parts. (An alternative interpretation would be that analysis of orientation is a more complex process than analysis of colour or size.)

The results are in some respects similar to those obtained in studies on selective listening. In both types of experiment subjects are able to select or to ignore items efficiently on the basis of their general physical characteristics, whereas selection appears to be difficult when it requires a more detailed identification of the items presented.

I wish to thank Mr. Pekka K. Lehtiö for help and stimulating discussion at all stages of the experiment. The investigation was supported by a grant from the National Research Council for the Humanities.

REFERENCES

AVERBACH, E., and SPERLING, G. (1961). Short-term storage of information in vision. In CHERRY, E. C. (Ed.) *Symposium on Information Theory*. London: Butterworth. Pp. 196–211.

BROADBENT, D. E. (1958). *Perception and Communication*. Oxford: Pergamon.

BROADBENT, D. E. (1963). Flow of information within the organism. *J. verb. Learn. verb. Behav.*, **2**, 34–9.

CHAPMAN, D. W. (1932). Relative effects of determinate and indeterminate Aufgaben. *Amer. J. Psychol.*, **44**, 163–74.

HARRIS, C. S., and HABER, R. N. (1963). Selective attention and coding in visual perception. *J. exp. Psychol.*, **65**, 328–33.

HESSELGREN, S. (1952). *Hesselgrens Färgatlas*. Stockholm: Palmer.

KÜLPE, O. (1904). Versuche über Abstraktion. *Ber. 1. Kongr. exp. Psychol.*, 56–68.

MELTON, A. W. (1963). Implications of short-term memory for a general theory of memory. *J. verb. Learn. verb. Behav.*, **2**, 1–21.

MOWBRAY, G. H. (1964). Perception and retention of verbal information presented during auditory shadowing. *J. acoust. Soc. Amer.*, **36**, 1459–64.

MURDOCK, B. B. (1966). Visual and auditory stores in short-term memory. *Quart. J. exp. Psychol.*, **18**, 206–11.

SPERLING, G. (1960). The information available in brief visual presentations. *Psychol. Monogr.*, **74**, No. 11.

SPERLING, G. (1963). A model for visual memory tasks. *Human Factors*, **5**, 19–31.

TREISMAN, A. M. (1964*a*). Monitoring and storage of irrelevant messages in selective attention. *J. verb. Learn. verb. Behav.*, **3**, 449–59.

TREISMAN, A. M. (1964*b*). Selective attention in man. *Brit. Med. Bull.*, **20**, 12–6.

Manuscript received 21st June, 1967.

SHORTER ARTICLES AND NOTES

THE INTERPRETATION OF THE PASSIVE VOICE

BY

P. N. JOHNSON-LAIRD

From the Department of Psychology, University College London

It was hypothesized that sentences in the passive voice emphasize the importance of the things referred to by their grammatical *subjects* to a greater extent than sentences in the active voice. Each subject had to produce simple diagrams to represent two sentences, one active and one passive, and it was assumed that the size of areas in these diagrams could be taken as an index of importance. In Group EQ, the sentences specified an equivalent arrangement of colours, e.g. "Red follows Blue," "Blue is followed by Red"; in Group CO, they specified converse arrangements, e.g. "Red follows Blue," "Red is followed by Blue." The predictions, that (i) the *subjects* of all sentences would tend to be represented as larger than the *objects*, and that (ii) the *subjects* of passives would be represented as larger than those of actives, were confirmed.

INTRODUCTION

Does the passive mean the same as the active in English? Some linguists claim that it does, e.g. Katz and Postal (1964), and others, e.g. Chomsky (1957), that it does not. McMahon (1963) and Slobin (1966) have shown that the passive is harder to understand than the active; and so, since the passive has not fallen into disuse, it is sensible to question whether it exists for merely stylistic purposes. The passive reverses the order of the grammatical *subject* and *object* of the active, and it also allows the active *subject* to be omitted altogether, e.g. "The plot was discovered." It seems, therefore, that one use of the passive is to emphasize the relative importance of whatever is referred to by its *subject*. The passive is chosen when its *subject* refers to a more predominant entity than its *object*, and this difference is sufficient to justify the presumably greater effort its production demands. It follows that the active is used when its *subject* predominates or when there is little difference in the importance of the entities denoted by *subject* and *object*.

If this hypothesis is true, then it will also apply to the way sentences are interpreted. The task in the present experiment was interpretative: the subjects had to illustrate two sentences by simple diagrams. On the assumption that the size of the areas representing *subject* and *object* provides an index of importance, it was predicted:—

(i) The *subjects* of the sentences would be represented by larger areas than their *objects*.

(ii) This difference would be greater for passives than for actives.

METHOD

Task. The subjects were given a slip of paper on which were drawn two long narrow rectangles, and their task was to illustrate two sentences by colouring the rectangles with crayon. Their attention was engaged by instructing them that "somebody else" would have to be able to match the diagrams with the appropriate sentences.

Design. Two Groups were used in order to vary the degree to which subjects would be likely to search explicitly for a difference in meaning between the two voices. Group EQ received an active and a passive sentence which referred to *equivalent* arrangements of colours:—

(1) Red follows Blue.

(2) Blue is followed by Red.

Group CO received sentences which referred to *converse* arrangements of colours.

(3) Red follows Blue.

(4) Red is followed by Blue.

In Group EQ, the task was intended to elicit a difference in meaning between the two voices: subjects would be forced to try to externalize the "rules" governing the usage of the active and passive. In Group CO, on the other hand, the task was intended to detect any implicit reliance upon such "rules." It was accordingly predicted (iii) that the difference in *subject* area between actives and passives would be greater in Group EQ.

Although each subject received only a single pair of sentences, these pairs were systematically varied to balance a number of factors. (*a*) In order to control for any directional effect inherent in the verb, half the pairs used the verb "follows" and the other half used "precedes." (*b*) All the pairs referred to arrangements of red and blue, but they were counterbalanced so that for each pair there was another pair which referred to the opposite arrangement of colours. (*c*) The order of the two sentences on the card was also balanced. Hence, there were eight different pairs for each of the two Groups, and subjects were allocated in rotation to them.

Procedure. Each pair of sentences was typed in capitals on a slip of paper (8 in. × ¾ in.), and the two rectangles, 5 in. × ¾ in., were drawn one beneath the other on another slip (8 in. × 2 in.).

The subjects sat at a table on which were the materials, including a red and blue crayon. They were given the following written instructions:—

> "There is in front of you a pair of sentences that you must regard as different though similar messages. Your task is to produce two drawings—by filling in the strips with crayon—so that the first drawing represents the first sentence and the second drawing represents the second sentence. Try to produce a pair of drawings such that, if they were given to somebody else, then he could correctly match the drawings with the sentences that they are meant to convey."
>
> "Each drawing must entirely fill in its strip—so do not leave any blanks; and the join between the two colours must be a sharp vertical line—so do not use any arrows, etc."

This second paragraph was introduced after pilot studies to eliminate some of the more bizarre solutions. Similarly, it was decided to allow subjects to have both strips in view throughout.

Subjects. The data were provided by 32 undergraduates (16 male and 16 female) in the Department of Psychology, University College London. They were all native-speakers of English and ignorant of the purpose of the experiment. (Five subjects failed to obey the instructions and were replaced by others.)

RESULTS

The mean lengths in inches of the areas representing the *subjects* of the sentences are given in Table I; the total length of each strip was 5 in. The results for the two different verbs were combined as the difference between them was insignificant. The measures involved in predictions (i) and (ii), though independent in sign, are not independent in magnitude.* Hence, nonparametric tests were used throughout.

TABLE I

MEAN LENGTHS IN INCHES OF *SUBJECT* AREAS

Groups	*Type of sentence*		*Overall*
	Active	*Passive*	
Group EQ..	3·0	3·5	3·2
Group CO..	2·7	3·0	2·8
Overall	2·8	3·2	3·0

---

* The author is indebted to R. J. Audley for this observation.

Table II shows the number of subjects whose mean *subject* length, computed from the active and passive drawings, was larger (or smaller) than mean *object* length. A Wilcoxon test on the combined data from both Groups was significant ($p < 0{\cdot}0005$, one-tail). Prediction (i), that the *subjects* of sentences would be represented by larger areas than the *objects*, was therefore confirmed.

TABLE II

NUMBER OF SUBJECTS WHOSE MEAN *SUBJECT* LENGTH WAS LARGER (OR SMALLER) THAN MEAN *OBJECT* LENGTH

Groups	*Larger*	*Smaller*	*Total*
Group EQ	11	5	16
Group CO	10	6	16
Total	21	11	32

Table III shows the number of subjects who made their passive *subject* larger (or smaller) than their active *subject*. There was one tie in CO. A Wilcoxon test on the combined data from both Groups was significant ($p < 0{\cdot}02$, one-tail). Prediction (ii), that the difference between *subjects* and *objects* would be enhanced for passives, was therefore confirmed.

TABLE III

NUMBER OF SUBJECTS WHOSE PASSIVE *SUBJECT* WAS LARGER (OR SMALLER) THAN ACTIVE *SUBJECT*

Groups	*Larger*	*Smaller*	*Total*
Group EQ	12	4	16
Group CO	9	6	15
Total	21	10	31

Although the trend was in the direction of prediction (iii), that the difference in *subject* area between actives and passives would be greater in EQ, a Mann-Witney test on the difference between the two Groups was not signfiicant.

It is feasible that those subjects who conformed to prediction (i) would conform to prediction (ii). Table IV is the appropriate contingency table derived from the combined data of both Groups. It is evident that there is some correlation, but it was not significant on a Kendall Rank Correlation test for two dichotomies.

TABLE IV

CONTINGENCY TABLE OF NUMBERS OF SUBJECTS CONFORMING (AND NOT CONFORMING) TO PREDICTIONS (i) AND (ii)

	Passive subject *larger*	*Active* subject *larger*	*Total*
Mean *subject* larger	16	4	20
Mean *object* larger	5	6	11
Total	21	10	31

There was a tendency, as can be seen in Table I, for *subjects* to be represented by larger areas in EQ than in CO. This was unexpected, but it was not significant on a Mann-Witney test.

Qualitative results

The "tactics" adopted by the subjects in EQ were relatively uniform. Of the 12 subjects who conformed to prediction (ii), (*a*) eight made the *subjects* of both statements larger than the *objects*, with passive *subjects* larger than those of the active; and (*b*) four divided the strip almost equally for the active but made the passive *subject* larger than its *object*. One subject adopting the latter tactic claimed that the passive reversed the direction implied by the active, and his diagrams resemble those of a subject in CO. Of the four subjects who did not conform to prediction (ii), two adopted tactic (*a*) and one adopted tactic (*b*), but in these cases the active was illustrated by the diagram appropriate to the passive, and the passive by the diagram appropriate to the active. The remaining subject illustrated the active by dividing the strip into two equal areas, and the passive by placing an area of the *subject* colour in the middle of the strip surrounded on both sides by the *object* colour.

The subjects in CO tended to represent the *subjects* of both sentences by slightly larger areas than the *objects*. Of the 10 subjects who did so, four conformed to prediction (ii), five did not conform, and one made the two areas equal. Another "tactic" was to divide both strips at the same place and to colour them appropriately. This yielded two out of three results conforming to prediction (ii). Of the remaining three subjects, whose results were all in accord with prediction (ii), two adopted strategy (*b*), and one represented both sentences by a series of vertical bands of colour, with more bands of the *subject* colour for the passive.

The majority of introspective reports were unrevealing. A few subjects in EQ came near to expressing the basic hypothesis.

DISCUSSION

The confirmation of the two main predictions supports the basic hypothesis, which is unaffected by the failure to confirm the subsidiary prediction. However, there are three points which must be considered. Firstly, the experimental sentences used only two verbs which stand in the relatively rare relation of denoting the converse of one another. The use of the passive to emphasize importance is perhaps less likely to occur when this emphasis could be achieved by using an alternative verb in the active voice; this would suggest that the use of these verbs makes an especially stringent test of the hypothesis. Secondly, the *subjects* of the pairs of sentences in EQ referred to different colours, whereas in CO they referred to the same colour. This imbalance is likely to have affected performance only in relation to the subsidiary prediction. It could have been controlled by using "mixed" pairs, i.e. one sentence with "follows" and the other with "precedes." But this was considered to be premature, since subjects might have exploited differences between the verbs themselves rather than between the two voices. Thirdly, the slight tendency for *subjects* to be represented by larger areas in EQ than in CO may have been due to a need to emphasize them in EQ in order to differentiate the diagrams. Any such tendency in CO would be merely an auxiliary aid, since the diagrams are differentiated by their opposite allocations of colour.

This experiment has shown that a passive sentence does not necessarily refer to exactly the same thing as its corresponding active. But this difference in reference must be mediated by a difference in meaning. If differences between *subject* and *object* areas can be taken as an index of predominance, then the basic hypothesis is substantially confirmed. Is the difference anything more than one of emphasis? Consider the sentence "Everyone knows some songs." This would normally be understood to mean that the songs were not necessarily the *same* ones for everyone. On the other hand, the sentence "Some songs are known by everyone" would normally mean that it *was* the same songs that were known. How are we to account for what Chomsky (1957) terms the "normal interpretation" of sentences like these? Clearly, it is due to word-order and in particular to the order of the two quantifiers "everyone" and "some." This order conveys which of the two quantifiers is the predominant one—which quantifier does the "binding" as the logicians term it. It is the effect of voice on word-order that is crucial rather than anything else, as is evident from the complementary examples, "Every song is known by someone" and "Someone knows every song."

The notion of predominance (or importance), ill-defined as it is, obviously involves the socio-physical setting and linguistic context of an utterance—factors which will have to be taken into account by any semantic theory. What is predominant depends on who you are, where you are, what you are doing, what you are talking about and to whom you are talking.

This research was performed under the supervision of Dr. P. C. Wason, to whom the author is extremely grateful for much encouragement, many stimulating ideas and criticisms, and a critical reading of an earlier draft of this paper. A report of the experiment formed part of the Ph.D. dissertation submitted to the University of London. An earlier version of this paper was presented to a meeting of the Experimental Psychology Society in London on 13th January, 1966.

REFERENCES

CHOMSKY, N. (1957). *Syntactic Structures.* The Hague: Mouton.
KATZ, J. J., and POSTAL, P. (1964). *An Integrated Theory of Linguistic Descriptions.* Cambridge, Mass.: M.I.T. Press.
McMAHON, L. E. (1963). *Grammatical Analysis as Part of Understanding a Sentence.* Unpublished Doctoral thesis, Harvard University.
SLOBIN, D. I. (1966). Grammatical transformations and sentence comprehension in childhood and adulthood. *J. verb. Learn. verb. Behav.*, **5**, 219–27.

Manuscript received 13th June, 1967.

INSTRUCTIONS, SELF-INSTRUCTIONS AND PERFORMANCE

BY

SHEILA JONES

From the Department of Psychology, University College London

Four semantically equivalent instructions were used for a classification task, in which response was required to the larger of two classes of items. The variables in the instructions were (1) the explicitness of the negative used to define the smaller class, and (2) the order of reference to this class within the sentence. Superior performance was obtained with instructions containing "except" compared with those using "not," regardless of order. Analysis of the self-instructions generated by the subjects, prior to performance, indicated that their form was largely determined by the given instruction. Speed of performance was related to form of self-instruction elicited prior to the task, but not to that elicited after completion of the task.

INTRODUCTION

In responding to an imperative sentence, the listener is only concerned with understanding what he has to do, the syntactic form of the sentence is relatively unimportant. In formulating his self-instruction, the listener may rearrange, replace, or omit parts of the original instruction. Access to the subject's self-instruction, immediately before he responds, should give information about his encoding of the sentence content. Asking the subject what he is going to do (rather than what he has been instructed to do), before the task is presented, should reveal the extent to which the original instruction has been retained. During performance of the task, the self-instruction may be modified to correspond more closely to the response requirements. Asking the subject what he has been doing, after the task is completed, should elicit his own description of the task and the extent to which his initial self-instruction has changed.

The present experiment seeks to determine the effect of syntactic variables in an instruction on the form of self-instructions and the extent to which these affect performance. A previous study by Jones (1966) of semantically equivalent instructions for a classification task influenced the choice of syntactic variables. An instruction of the form "Respond to all items except x, y, z," was found to produce a large number of errors due to incorrect response to items x, y, z. It was hypothesized that a more overt reference to the negative response required to x, y, z, might reduce these errors. The alternative form of instruction "Do not respond to items x, y, z, respond to all the rest," as well as making the negative more explicit (by replacing "except" by "not") also reverses the order of reference to the negative class in the sentence. Rothkopf (1963) has shown that varying the order of clauses in a sentence affects recall and acquisition. It was therefore decided to investigate the effect of the two variables: order of reference to the negative class within the sentence (first or second), and explicitness of negative "except" or "not." In addition to the two forms of instruction already referred to, the two less common forms "Except for items x, y, z, respond to all the rest," and "Respond to all items but not x, y, z," were included. These four semantically equivalent instructions were used to investigate the effect of syntactic variables on self instructions and performance.

METHOD

Design. Sixty-four subjects (32 men, 32 women) all undergraduate or research students were given a cancelling task in which the same digits had to be marked by all subjects. The subjects were assigned in rotation to one of four types of instruction, denoted by N_1, N_2, IN_1, IN_2 respectively, according to whether the negative used was "not" (N) or the implicit negative "except" (IN), and whether the negative class was referred to first (1) or second (2) in the instruction. The instructions were:

Instruction Type	N_1	Do not mark the numbers 2, 5, 8, mark all the rest.
	N_2	Mark all the numbers but not 2, 5, 8.
	IN_1	Except for the numbers 2, 5, 8, mark all the rest.
	IN_2	Mark all the numbers except 2, 5, 8.

The digits used were 1, 2, 3, 4, 5, 6, 7, 8 and these occurred with equal frequency and were arranged in different random order on each trial. The subjects were given no information about the range or relative frequency of the digits.

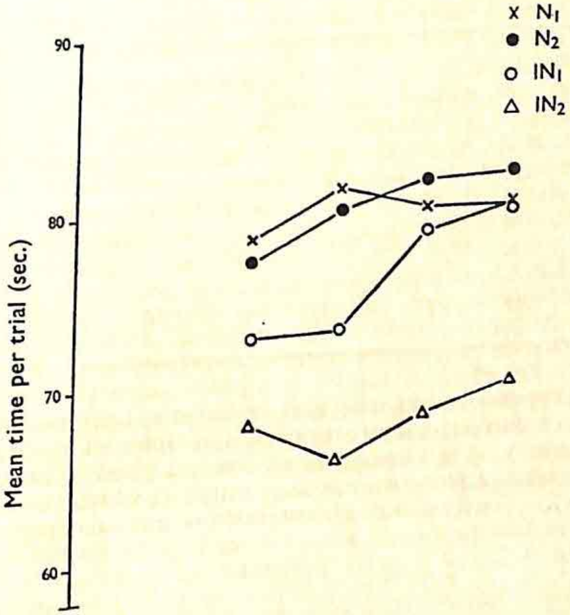
Materials. The same material was used by all subjects and consisted of four pages of digits, each page containing 12 rows of equally spaced digits (144 digits per page).

Procedure. Each subject was tested individually and was asked to work as quickly as possible, but at the same time to make as few errors as possible. A sample page of digits was shown to the subject before the experiment began (and before the instruction was given). A demonstration was given of the method of cancelling to be used (a stroke through each digit). The subject was told that he would be timed, and was asked not to correct errors. One of the four forms of instruction was then given to the subject, written on a card. He was told it would be removed before the task was presented. When he indicated he was ready, the instruction was removed. The subject was then asked "Tell me what you are going to do?" His response was noted. The task was then presented. After completing four trials, which were each timed, the subject was asked "Tell me what you have been doing?" His response was again noted.

RESULTS

Effect of instructions on performance. The mean response times in seconds per trial for each form of instruction are shown in Figure 1. Fastest performance was obtained by subjects with instruction IN_2 on all four trials. An analysis of variance was carried out on the total times over four trials, in terms of a 2 × 2 factorial design, with explicitness of negative and order of reference to the negative class (first or second within the sentence), as the two factors. Explicitness of negative was a significant factor, i.e. the implicit

FIGURE 1



Speed of performance as a function of form of instructions for a digit cancelling task.

Instruction N_1: Do not mark the numbers 2, 5, 8, mark all the rest.

N_2: Mark all the numbers but not 2, 5, 8.

IN_1: Except for the numbers 2, 5, 8, mark all the rest.

IN_2: Mark all the numbers except 2, 5, 8.

forms IN_1 and IN_2 gave faster performance ($F = 5{\cdot}42$, $d.f. = 1, 60$, $p = 0{\cdot}025$) than the explicit forms N_1 and N_2. Order of reference to the negative class within the sentence was not a significant factor. The chief comparison of interest was between instructions (IN_2), "Mark all the numbers except 2, 5, 8" and (N_1) "Do not mark the numbers 2, 5, 8, mark all the rest." These were the two forms initially suggested as commonly used alternatives for this type of task. Orthogonal comparison of overall time scores showed significantly faster performance under form IN_2 than N_1. ($F = 6{\cdot}11$, $d.f.\ 1, 60$, $p = 0{\cdot}025$).

Errors

Errors made on the positive class of digits (omissions) were analysed separately from those made by responding incorrectly to the negative class of digits (false positives).

Omissions. The percentage of errors on the positive class of digits was 0·24 for instruction type IN_1, 0·28 for form IN_2, 0·19 for form N_1 and 0·36 for form N_2.

False positives. The percentage of errors on the negative class of digits was 4·23 for instruction form IN_1, 3·65 for form IN_2, 3·56 for form N_1, and 5·64 for form N_2.

Comparisons between percentage of errors for different instruction forms revealed no significant differences for either type of error.

Initial self-instructions (*prior to performance*). Verbal responses to the question "What are you going to do?" obtained from subjects, immediately before the task was presented, were analysed in relation to the given instructions. The self-instructions fell into three main types (*a*) those closely resembling the given instruction ($n = 23$), (*b*) those resembling one of the other three forms ($n = 23$) and (*c*) those in which reference was made only to the negative class of digits ($n = 18$). (Of the latter, 12 were abbreviated forms, i.e. "not 2, 5, 8" or "except 2, 5, 8" and six omitted the negative prefix altogether.)

TABLE I

Relationship between Instructions and Initial Self-instructions ($n = 64$)

		Initial self-instruction type				
		IN_1	IN_2	N_1	N_2	n
Instruction type	IN_1	10	4	2	0	16
	IN_2	3	13	0	0	16
	N_1	2	4	10	0	16
	N_2	0	8	5	3	16
	n	15	29	17	3	64 (N)

Table I shows the types of self-instruction in relation to the given forms, i.e. categorized according to degree of explicitness of negative and order of reference to the negative class. Abbreviated forms of self-instructions were classified as N_1 or IN_1 according to whether the negative was explicit or implicit. Those in which the negative was omitted were classified as IN_1. Nearly half the sample ($n = 29$) used the self-instruction form IN_2 and this form of instruction was the one retained by the largest proportion of subjects (13 out of 16). Forms IN_1 and N_1 were both retained by 10 out of 16 subjects, but only 3 subjects retained the form N_2. (The fact that form N_2 was not chosen spontaneously by any subject as self-instruction suggests that it would rarely be used in this type of task.)

Subjects with instructions containing "except" tended to retain the implicit negative in their self-instruction. Subjects with instructions containing "not," however, showed a marked tendency to replace it with an implicit negative in their self-instruction. Three-quarters of the sample retained the same order of reference to the negative class as given in their instruction, and the overall change in order was nil.

A significant number of subjects changed their given instruction to a self-instruction of the form IN_2. (McNemar test for significance of changes: $z = 2{\cdot}75$, $p = 0{\cdot}006$.) Thus the change to implicit negatives is a reflection of a preference for the use of "all except . . ." rather than a substitution of "except" for "not."

Initial self-instructions and speed of performance. Time scores of subjects were reclassified according to the type of self-instruction chosen. (The 3 subjects using form N_2 were omitted from this analysis as no subject chose this form spontaneously.) Within each of the three main forms of self-instruction, IN_1 ($n = 15$), IN_2 ($n = 29$) and N_1 ($n = 17$), there was no significant difference in time scores between those subjects *retaining* their given form of instruction and those who *changed* their given forms. This suggests that although the form of instruction influenced the *choice* of self-instruction, subsequent speed of performance was determined by the type of self-instruction used.

Average overall speeds under the three types of self-instruction were 258·8 sec for form IN_2, 305·1 sec for form IN_1, and 340·3 sec for form N_1. These speeds correspond in order of magnitude to those found for instructions of the same form (Fig. 1). Comparison of time scores for subjects using forms IN_1 and N_1 enabled the effect of explicitness of negative to be determined directly. No significant difference in speed was found. A comparison was made between the combined time scores for self-instruction forms IN_1 and N_1 with those for the preferred form of self-instruction IN_2. Speed of performance was significantly faster under form IN_2 (Mann-Whitney, $z = 2·16$, $p = 0·03$).

Final self-instructions (after performance of task). The responses from subjects, after completing the task, to the question "What have you been doing?", were analysed. (This question was put only to the final 48 subjects of the sample.) The majority of responses (43 out of 48) were complete sentences, the remainder being abbreviated forms, similar to the types of self-instruction elicited before the task was presented.

TABLE II

RELATIONSHIP BETWEEN INITIAL AND FINAL SELF-INSTRUCTIONS (n = 48)

		Final self-instruction type				
		IN_1	IN_2	N_1	N_2	n
Initial self-instruction type	IN_1	2	5	1	1	9
	IN_2	1	19	3	0	23
	N_1	2	6	5	1	14
	N_2	0	2	0	0	2
	n	5	32	9	2	48 (N)

Table II shows the relationship between initial and final forms of self-instruction. Two-thirds of the subjects ($n = 32$) gave final instructions of form IN_2. Comparison with the subjects' initial self-instructions showed a significant change to this form after performance (McNemar test: $z = 1·94$, $p = 0·026$). Time scores for the 48 subjects were re-analysed in relation to the three main types of final self-instruction. The average overall speeds were 293·4 sec. for form IN_2, 305·0 sec. for form IN_1, and 322·8 sec. for form N_1. Comparisons of scores in terms of explicitness of negative and order of reference to the negative class showed no significant differences. There appeared to be no relationship between speed of performance and final form of self-instruction.

DISCUSSION

The finding that the self-instruction most frequently elicited prior to the task coincided with the optimal one for efficient performance may be related to the particular experimental conditions used. Linguistic habit (Miller, 1964) could account for the preference for the form "Respond to all the items except x, y, z," in the present context. This self-instruction implies a division into two classes—the larger class consisting of positive items. The majority of the other self-instructions were prohibitives referring only to the negative class of items, e.g. "Do not respond to x, y, z." Eifermann (1961) has shown that prohibitive connotations associated with "not" contribute to response latency. This would account for the reduction in speed obtained with the more prohibitive type of self-instructions.

Final self-instructions, elicited after completion of the task, showed an increased preference for the form "Respond to all items except x, y, z." Almost all the final self-instructions were complete sentences and this, together with the finding that the forms chosen were unrelated to previous performance, suggested that they were merely preferred forms of descriptions of the task. In other contexts, the present results suggest that self-instructions elicited *prior* to performance may reveal linguistic factors which affect performance, whereas those elicited *after* performance may indicate preferred descriptions of the task not necessarily related to performance.

REFERENCES

EIFERMANN, R. R. (1961). Negation: a linguistic variable. *Acta Psychol.*, **18**, 258–73.

JONES, S. (1966). The effect of a negative qualifier in an instruction. *J. verb. Learn. verb. Behav.*, **5**, 495–501.

MILLER, G. A. (1964). Language and Psychology. In LENNEBERG, E. H. (Ed), *New Directions in the Study of Language*. M.I.T. Press, 89–107.

ROTHKOPF, E. Z. (1963). Learning from written sentences: within-sentence order in the acquisition of name clause equivalences. *J. verb. Learn. verb. Behav.*, **2**, 470–5.

Manuscript received 16th June, 1967.

RESPONSE LATENCIES IN SHORT-TERM MEMORY

BY

BENNET B. MURDOCK, JR.
From the Department of Psychology, University of Toronto

This paper deals with the problem of the interpretation of response latencies in short-term memory. A paired-associate study using a probe technique was conducted with the main experimental variation the length of the recall interval (1, 2, or 4 sec.). While shortening the interval had a statistically significant effect on recall probability the interaction between recall interval and probe position was negligible. While traditionally response latency is considered a measure of associative strength, such an interpretation seems inappropriate here. As an alternative, latencies may reflect more criterion values than sensitivity as these measures are interpreted in signal-detection theory.

INTRODUCTION

It seems to be generally assumed (e.g. Eimas and Zeaman, 1963; Williams, 1962) that, in memory, response latencies (or their reciprocal, speed) are a measure of associative strength. In fact, in the first reference above it was implied that latencies were an even more sensitive measure of associative strength than were response probabilities. Such a view is perhaps one of our legacies from Hullian behaviour theory, wherein it was formally stated (Hull, 1943, p. 104) that reaction latency is one of the various ways in which habit strength may manifest itself. And, of course, in word-association work this particular relationship is referred to as Marbe's Law (Woodworth and Schlosberg, 1954).

A recent study (Murdock, 1963*b*) using paired associates and a probe technique yielded some rather surprising results: curtailing the recall interval had a greater (adverse) effect on the last pair in the list than on the earlier pairs. This finding was surprising because, with unlimited recall intervals, recall latencies of early pairs were markedly longer and recall probabilities lower than those of later pairs. In the unrestricted case latencies and probabilities are generally correlated (longer latencies associated with lower probability scores), but the results cited above suggest that this correlation might be at least partly fortuitious.

More particularly, in short-term memory (STM) response latencies may reflect how much time a subject *chooses* to take to respond rather than how much time he *must* take to respond. In terms of signal-detection theory (TSD) it might be that response latencies in a probe task reflect more criterion factors than sensitivity or associative "strength" *per se*. A recent study found that, "Within any given retention interval latencies and confidence judgements are highly correlated.... That confidence judgements and latencies may be interchangeable is suggested by the fact that symmetric ROC curves are obtained when one partitions on the basis of latency . . . rather than partitioning on the basis of confidence judgements" (Murdock, 1966, p. 322). In this study d′ did not change over probe position though recall probability clearly did. Thus, if "strength of a response" is properly assessed by the d′ measure of TSD, then in this study latency of response was associated with criterion factors rather than sensitivity.

The present experiment represents a further study of this problem. Essentially it extends the Murdock (1963) study in that (*a*) only one probe test was given on each trial, and (*b*) the recall period was systematically varied over a 4:1 range. Point (*a*) is important in that the previous results may have reflected the cascaded output tests; performance on the first probe (the only one considered here) may have been influenced by other probes yet to come. Point (*b*) reflects the view that the most frutiful way to study this problem is to impose experimentally-determined limits on task performance (in this case essentially by means of a rather extreme payoff matrix) rather than try to find out what the subject may do under less structured conditions.

METHOD

Materials. The stimulus material for the lists of paired associates consisted of common English words paired at random. All words were two syllables and not more than eight letters long; they were randomly selected from the Thorndike-Lorge (1944) G count of 20 and above except that homophones, contractions, and archaic words were excluded.

Procedure. Each list consisted of six A-B pairs followed by a single probe, the A member of one of these six pairs. When the probe was presented the subject had 1, 2, or 4 sec. in which to respond (i.e. with the appropriate B member); this experimental manipulation constituted the main variable of interest. A voice-key arrangement was devised which was activated upon presentation of the probe and started a Standard Electric Timer with a 1-sec. sweep hand. The subject had to respond by speaking the correct response into a microphone so as to stop the timer within the allotted time (1, 2, or 4 sec. according to condition). The clock was clearly visible and credit for a correct response was given only when the deadline was met. There was little difficulty either with this procedure or with this apparatus.

For greater generality two modes of presentation were used, visual and auditory. For visual, the words were photographed on 16 mm. film one word per frame and projected by means of a Dunning Animatic single-frame projector. For a presentation time of 2 sec. per pair the projector was stepped at a rate of 120 frames per minute; the two words of each A-B pair were photographed on successive frames which were followed by two blank frames as a study period. For auditory presentation the same lists were recorded on an Ampex F-44 Fine Line tape recorder so as to make the visual and auditory as similar as possible. That is, the two words of each pair were spoken at a rate of 120 words per minute, and each pair was followed by a 1-sec. quiet period as in the visual case.

The probe was presented in the same mode as the list; for auditory, it was recorded on the tape and for visual it was photographed on the third frame following the last pair (to allow two blanks for the last pair). During presentation the experimenter monitored the display and started the Standard Electric Timer upon presentation of the probe.

Subjects. The subjects were 24 students from the Toronto subject pool, 12 of each sex All were tested individually.

Design. The design was an eight-way factorial, $2 \times 2 \times 3 \times 2 \times 2 \times 6 \times 3 \times 2$ The first four variables were between-subjects variables which provided the necessary counterbalancing controls and assigned each subject to a unique condition. Variable A controlled for mode order; half the subjects went from visual to auditory while the other half went from auditory to visual. Variable B made orthogonal lists and mode; all lists that were visual under one level were auditory under the other level, and vice versa. Variable C controlled against differential practice effects biasing the main experimental comparison; response time (1, 2, or 4 sec.) cycled every six lists and three different orders were devised following customary Latin-Square principles. Variable D was sex.

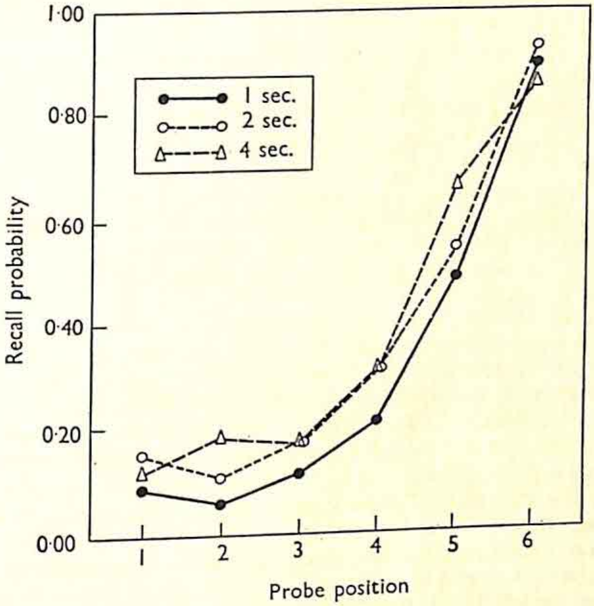
The last four variables were within-subjects variables which provided the 72 lists for each subject. Variables E, F, G, and H were Mode, Probe, Time, and Replication, respectively, with Mode either auditory or visual, Probe one of the six serial positions, Time either 1, 2, or 4 sec., and two replications. With this design every single response by every single subject is assigned to a unique cell in the eight-way classification, and all experimental comparisons of interest are within-subjects effects for greater sensitivity.

RESULTS

The main results are shown in Figure 1, where probability of recall is shown as a function of probe position separately for response times of 1, 2, and 4 sec. The auditory and visual conditions have been pooled in this graph because the interaction of mode and response time was non-significant ($p > 0{\cdot}05$). It is abundantly clear that the effect of response time was (*a*) rather small, and (*b*) relatively constant over all probe positions.

To insure against artefacts of averaging, several analyses of variances were performed, collapsing over the between-subjects variables in different ways (specifically, first over A and B, then over A, B, and C, and finally over A, B, C, and D), in all cases with an arcsin transformation. The results were consistent; as expected mode, probe, and their interaction were all clearly significant. More pertinent, however, is that the effects of response time were in fact statistically significant ($F(2, \infty) = 6{\cdot}01$, $p < 0{\cdot}05$) but the interaction of response time with probe position was clearly not ($F < 1{\cdot}0$).

FIGURE 1



Recall probability as a function of probe position for the three recall intervals.

Discussion

For unlimited recall periods, median response latencies of all correct responses in a previous study (Exp. IV, Murdock, 1963*a*) were found to be 0·78, 1·01, 1·50, and 1·78 sec. for probes 6, 5, 4, 3–1 respectively. The proportion of responses from each distribution which exceeded 4·0 sec. were 0·003, 0·079, 0·135, and 0·159 (again respectively) for these four probe positions. Since probability of correct recall showed a commensurate change, there would be every reason to suppose that in fact probability and latency are interchangeable measures of response strength; this is the point stated in the Introduction.

The present experiment, however, suggests that this conclusion is untenable. The main reason (and the chief point of the present experiment) is that, as shown both in Figure 1 and in the analyses of variance, the interaction between response time and probe position was negligible. Stated otherwise, response time did (statistically at least) exert a significant effect but the effect was no different for early or late pairs in the list. In fact the greatest difference seemed to occur at Probe 5, the penultimate pair in the list. But were response latencies really an indication of the longer time needed for weaker response to assert themselves then clearly the effect of limiting this response time should have been greater for the early probe positions.

In a recent study of reaction times to weak signals in noise Carterette, Friedman, and Cosmides (1965) found that latencies of "yes" response consistently decreased as *a priori* (signal) probability increased, and that in general the latency distributions resembled each other according to the response ("yes" or "no") rather than according to the stimulus (signal or noise alone). Exactly comparable results have been found in a previous study of paired associates in STM (Murdock, 1966). Elsewhere (Murdock, 1967) a simple stochastic model has been suggested wherein recall probability drops off as a function of input and/or output interference. Given then that recall probability is a function of number of intervening pairs, then the ideal observer of TSD must adjust his criterion accordingly. The criterion for early pairs becomes more strict and so latency increases; not because these responses are "weaker," but because the criterion level is higher.

A rationale for this presumed criterion shift comes from TSD. Consider a payoff matrix where the observer is rewarded for hits and correct rejections but penalized for misses and false alarms. Then, if this payoff matrix is symmetric, it is easy to show that, to maximize expected value, the observer should determine his criterion solely on the basis of the *a priori* probabilities (specifically, the ratio of the ordinates of the overlapping distributions should be equal to the odds against the occurrence of a signal). In the

STM situation, it is suggested that the interference (input and/or output) determines the probability values, and that the subject follows a maximum-likelihood principle in that his criterion is higher for pairs tested at longer retention intervals (where, of course, probability is lower). Response latencies, then, reflect more the criterial level than the strength of the association. While the present data do not demand this particular interpretation, they are certainly consistent with it.

This research was supported by Grant APA 146 from the National Research Council of Canada and by Grant GB 4545 from the National Science Foundation. The writer is grateful to Muriel Mougenal for experimental assistance and to Wayne Donaldson for a helpful reading of the manuscript.

REFERENCES

CARTERETTE, E. C., FRIEDMAN, M. P., and COSMIDES, R. (1965). Reaction-time distributions in the detection of weak signals in noise. *J. acoust. Soc. Amer.*, **38**, 531–42.

EIMAS, P. D., and ZEAMAN, D. (1963). Response speed changes in an Estes' paired-associate "miniature" experiment. *J. verb. Learn. verb. Behav.*, **1**, 384–8.

HULL, C. L. (1943). *Principles of Behaviour.* New York: Appleton-Century.

MURDOCK, B. B., Jr. (1963*a*). Short-term retention of single paired associates. *J. exp. Psychol.*, **65**, 433–43.

MURDOCK, B. B., Jr. (1963*b*). Interpolated recall in short-term memory. *J. exp. Psychol.*, **66**, 525–32.

MURDOCK, B. B. Jr. (1966). The criterion problem in short-term memory. *J. exp. Psychol.*, **72**, 317–24.

MURDOCK, B. B., Jr. (1967). A fixed-point model for short-term memory. *J. math. Psychol.*, **4**, 501–6.

THORNDIKE, E. L., and LORGE, I. (1944). *The Teacher's Word Book of 30,000 Words.* New York: Teachers College, Columbia Univ.

WILLIAMS, J. P. (1962). A test of the all-or-none hypothesis for verbal learning. *J. exp. Psychol.*, **64**, 158–65.

WOODWORTH, R. S., and SCHLOSBERG, H. (1954). *Experimental Psychology.* New York: Holt.

Manuscript received 12th August, 1967.

ACOUSTIC FACTORS IN VISUAL SEARCH

BY

D. W. J. CORCORAN* and D. L. WEENING

From the U.S. Navy Electronics Laboratory Center, San Diego, California and San Diego State College

Previous work has shown that in searching for existing or absent "e.s" in printed prose, the presence or absence of silent "e.s" was less likely to be detected than that of pronounced "e.s." It was suggested that the acoustic or kinaesthetic "image" was searched for evidence of an "e" in addition to the visual stimulus and that evidence from both sources was considered in making the appropriate response.

The present experiment employs mainly substitutive errors within words, which may or may not change their pronunciation. The results suggest that the *form* of the acoustic correlates has no bearing upon whether the words are detected as wrongly spelt, but that the presence or absence of an acoustic event corresponding in time to the spatial location of the error is important.

INTRODUCTION

Two previous studies (Corcoran, 1966, 1967) have investigated the role of acoustic factors in the search for existing or absent letters in words contained in passages of written prose. In the first study subjects were presented with a passage taken from *Punch* and required to cancel all the letters "e" as quickly as possible. The main finding was that "e.s" which were normally silent were less likely to be detected than pronounced "e.s." The second experiment showed that in searching for wrongly spelt words in an editorial of a national newspaper, those words whose missing "e" would, if present, have been silent were less likely to be noticed as errors than the words in which the missing "e" was normally pronounced. These studies suggested that, in the process of searching the words, verbalization occurred and the resulting acoustic or kinaesthetic "image" was scrutinized for evidence of an existing or absent "e." The detection of the presence or absence of an "e" was therefore based upon evidence from two sources.

In the first study a comparison was also made between the probabilities of detecting "e.s" pronounced ɜ as in w*e*re, ɛ as in m*e*t and /i/ as in *e*ve. No differences were evident and it was tentatively concluded that the nature or quality of the acoustic correlate of the visual stimulus was of no consequence, only the fact that an acoustic event was or was not present to correspond with the visual stimulus. However, the frequency with which "e" is pronounced ɜ, ɛ or /i/ may not be very different and a more exact test of whether the form of the acoustic event influences visual detection is therefore necessary. The following experiment was conducted to determine (1) whether the quality of the acoustic image did or did not affect the decision as to whether a word was incorrectly spelt, and (2) whether the silent-pronounced phenomenon was specific to letters "e."

METHOD

Material, subjects and procedure

A passage of 663 words was selected from Anne Morrow Lindberg's *Gift From the Sea.* Several changes were made in order to balance the following six categories of error:—

(1) Twelve letters "c" were substituted for "s" (c–s).
(2) Twelve letters "s" were substituted for "c.s" which were pronounced /s/ (s–c/s/).
(3) Twelve letters "s" were substituted for "c.s" pronounced as /k/ (s–c/k/).
(4) Three silent "k.s" were omitted (k/ /).
(5) Three pronounced "k.s" were omitted (k/k/).
(6) Twelve random substitutive errors were incorporated (r).

* Exchange scientist from the M.R.C. Applied Psychology Research Unit, Cambridge, England.

The errors incorporated in categories (1), (2) and (3) were cross matched such that errors occurred at approximately comparable places in words of approximately comparable length whose probability of occurrence in language was also controlled to some extent (Thorndike and Lorge, 1944). All silent "k.s" were located at the beginning of words and words beginning with "k" of comparable length were used as controls under category (5). As a result of a typing and an experimental error two words were changed, which resulted in correctly spelled words which were nonsense in the particular context. These are included here because of their interesting consequences. Apart from the two errors none of the changes of spelling resulted in a correctly spelled word of different meaning.

Thirteen students (seven males and six females) of San Diego State College served as subjects. They were required to underline all the words in the passage, which they considered wrongly spelt and asked to work as quickly as possible. The slowest worker took 5 min. 25 sec. to complete the task, the fastest took 2 min. 22 sec. and the median time was 3 min. 25 sec. After the task had been completed subjects were given an oral test of spelling in which the words they had failed to detect as errors were used as items. In the single instance in which an error in the main task coincided with spelling error, the datum was disregarded in the main analysis.

Results and Discussion

Table I contains the probability of an error under each of the six categories. Only one relevant comparison in this table was statistically reliable; this was the difference between the probabilities of detecting silent and pronounced "k.s," (4) vs. (5) ($t = 2{\cdot}29$, $p < 0{\cdot}05$). One subject failed to detect two errors under category (4) and four others failed in single instances.

The typing error consisted of omitting the "k" in "know"; this error was not detected by any of the subjects despite the fact that in the sentence "The life I have chosen as wife and mother I (k) now to involve . . ." it was nonsense. The experimenters' error was the substitution of s for c in "devices." Both the legal meaning and the third person of the verb devise is rare, yet it resulted in five errors although it was clearly out of context.

The substitution of "s" for c/s/ can be considered to produce a "word" which is phonetically less changed by the substitution than one produced under categories (1), (3) and (6). Thus, if the *quality* of the acoustic correlate influences the detection of a wrong letter, more errors should have occurred under category (2) since evidence of the existence of an error from the acoustic source would presumably be lacking. This was not evident

TABLE I

List of Changed Words and Number of Failures to Detect the Error

	(1) c–s		(2) s–c/s/		(3) s–c/k/		(4) k/ /		(5) k/k/		(6) r	
	situation	3	balancing	3	America	3	(k)nitting	1	(k)eeping	0	even (j)	3
	compromise	2	crevices	3	unification	2	(k)nack	3	(k)icks	0	more (t)	3
	parasol	2	significance	2	physical	2	(k)nots	2	(k)itchen	0	with (p)	3
	choose	2	electricity	1	political	2					sing (l)	2
	transport	1	replace	1	intellect	1					equipment (o)	2
	utmost	0	grace	1	carriage	1					marketing (p)	0
	thousand	0	cynical	1	conclude	1					husband (m)	0
	socialised	0	centralised	0	camps	1					puberty (o)	0
	suburbs	0	civilised	0	facts	0					butcher (l)	0
	soulful	0	certain	0	countless	0					thorough (r)	0
	sewing	0	cycle	0	contact	0					women (k)	0
	stove	0	cease	0	carry	0					five (p)	
Probability of failure to detect error	0·06		0·08		0·08		0·15		0·00		0·08	

The letters in parentheses in the r column indicate the substituted letter.

in the data. However, the missing silent "k.s" were not well detected, which supports the findings for letters "e" in the previous studies. It would appear therefore that the earlier conclusions were probably valid in suggesting that the form of the acoustic event is of little consequence, provided that a "phoneme" of some kind exists in the acoustic image corresponding to its spatial position within the visual stimulus.

It can be seen clearly in Table I that detections were not evenly spread over the 54 textual errors. The greater majority of the words which were always detected as errors contained a substitution of the initial letter. In the cancellation experiment (1966) it was noted that the earlier the "e" occurred in a word the *higher* was the probability of its being detected. In the 1967 experiment, however, in which "e.s" were purposely *omitted*, the absent "e.s" located at an early position in a word were *less* likely to be detected. Thus the early position within the words are apparently well scanned for the presence of events such as "e.s" and substitutive errors, but poorly scanned for the absence of "e.s" and silent "k.s."

REFERENCES

CORCORAN, D. W. J. (1966). An acoustic factor in letter cancellation, *Nature*, **210**, 658.

CORCORAN, D. W. J. (1967). Acoustic factors in proof reading, *Nature*, **214**, 851.

THORNDIKE, E. L., and LORGE, I. (1944). *The Teacher's Word Book of* 30,000 *Words*, New York: Columbia University Press.

Manuscript received 8*th September*, 1967.

COMMENT ON "SPONTANEOUS ALTERNATION AND PATTERN OF REINFORCEMENT"

BY

DAVID LESTER

From the Department of Psychology, Brandeis University, Waltham, Massachusetts, U.S.A.

FURTHER DISCUSSION

BY

A. W. STILL

From the Department of Psychology, University of Durham

The aim of this experiment was to replicate the phenomenon reported by Still that the spontaneous alternation of non-rewarded rats declines faster over repeated testing than that of rewarded rats. The present experiment failed to find such an effect. It is suggested that on the basis of these experiments, it would be unwise to attribute any effect on spontaneous alternation behaviour with repeated testing to the rewarding or non-rewarding of rats.

INTRODUCTION

Still (1966) has suggested that spontaneous alternation is due to an exploratory tendency, but that after prolonged testing the behaviour disappears. However, if the animal is rewarded in the situation on every trial, then the spontaneous alternation behaviour does not disappear since the animal has been rewarded for alternating. Still's evidence for this was an experiment with five rats in one group and four rats in another group, in which each animal was given two trials a day in a T-maze when deprived. One group was rewarded; the other group was not rewarded. After 24 days of testing the rate of spontaneous alternation of the non-rewarded rats dropped below that of the rewarded rats. This effect has not been reported before, but previous experiments have used only short periods of testing. For example, Spear, Hill, and O'Sullivan (1963) used only 6 days of testing. Freedman (1965) found that over 40 days of testing stimulus alternation in rewarded rats declined significantly, but he did not include a non-rewarded group of animals.

Before we accept Still's explanation, or one that would view reward as increasing a general drive state (including a hypothetical exploratory drive), we must be sure that the phenomenon exists and is not an artefact. Two possible artefacts exist. First, the rewarded rats may alternate on trial 2 because they have eaten all the food in the arm visited on trial 1. Spear, Hill, and Campbell (1964) found that this factor had no effect on the alternation rate of rats in a 6-day testing period, in an experiment in which one group of rats was allowed to finish the food allotment each trial and another group was not allowed to finish their allotment of food.

An alternative explanation is based on the fact that non-rewarded rats run more slowly and so may indulge in more exploration of the approaches to the arms and make more partial entries than rewarded rats. In this case the use of a complete body entry into an arm, while making good experimental sense does not make good psychological sense. The non-rewarded rats used by Still may have been alternating their initial partial exploration of the arms, but this would not show up when a criterion of a full body entry is used.

The aim of the present experiment was to replicate Still's experiment with a larger number of rats. Further, the decision times of the rats were taken and examined to see if a decline in alternation was accompanied by a lengthening of these decision times.

METHOD

Subjects

The subjects were 17 male and 12 female albino rats of the Charles River Breeding Laboratory strain CD. They were the product of two litters. They were thoroughly tame. (No animal defecated or urinated on any trial during the course of the experiment.) The rats were 64–68 days old at the start of the experiment.

Apparatus

The apparatus consisted of a wooden T-maze covered with wire-netting. The arms and stem were 22 in. long, 6 in. high, and 6 in. wide. There were guillotine doors at the entrances to each arm.

Procedure

On day 1 the rats were put on a feeding schedule in which they were fed for 4 hr. each afternoon. On days 1–4 the rats were left alone except for feeding. On days 5–7 each rat was put in one arm of the T-maze for 5 min. Half found Frosty-O cereal in the arm (a General Mills Food Product). All animals received their daily food allowance after this exposure. The rewarded group of animals comprised six females and eight males. The non-rewarded group comprised six females and nine males.

On days 8 to 37 the rats were tested in the T-maze. Each rat was given two trials a day with a 5-sec. intertrial interval. On days 8–12 each rat was detained in the arm chosen for 60 sec. and the rewarded rats received three Frosty-Os. On days 13–14 the detention time was reduced to 30 sec. and the reward was reduced to two Frosty-Os. On days 15–37 the detention time was reduced to 15 sec. and the reward was reduced to one Frosty-O (similar to Still's condition). The rats were introduced to the final experience gradually in order that the rewarded rats should eat the reward on all trials. On the first day of testing only two rats failed to eat and on the second day only one rat. Thereafter all rewarded rats ate on all trials. The proportion of rewarded rats eating on the early days of testing would not have been so high had the detention time been only 15 sec. and the reward only one Frosty-O.

The time from when the rat was placed into the T-maze until it had entered one arm completely was timed with a stop-watch to the nearest 0·1 sec. An entry was defined as made when the whole body of the animal had entered the arm and the guillotine door could be closed. On occasions when the rat escaped before the door was closed, the rat was placed into the arm entered, but the information from that trial was discarded if the animal escaped on the first trial of the day.

RESULTS

Alternation behaviour

The percentage of alternation for all 30 days of testing was 92·6 per cent. for the non-rewarded rats and 93·3 per cent. for the rewarded rats. These percentages did not differ significantly.

There were no trends in the alternation behaviour of the two groups of animals over the 30 days of testing. For the first 3 days of testing the non-rewarded rats alternated on 91·1 per cent. of the trials and the rewarded rats on 90·5 per cent. For the last 3 days of testing the non-rewarded rats alternated on 95·5 per cent. of the trials and the rewarded rats on 81·0 per cent. On the final day all of the 15 non-rewarded rats alternated and 11 of the 14 rewarded rats alternated. This difference has a Fisher exact probability of occurrence of 0·10. Little confidence should be put in the validity of this difference on the final 3 days of testing, for the difference was in the opposite direction for the block of three trials before the final block.

The percentages of alternation for blocks of 6 days remained between 90·5 and 95·3 per cent. for all rats. The percentages are shown in Table I.

Running times

The mean running times for the non-rewarded rats were greater than those for the rewarded rats on all days of testing except the first day. The binomial probability for this (29 out of 30) is less than 0·001.

The difference between the groups first reached significance at the 0·05 level of significance on the third day of testing ($t = 2{\cdot}24$, $d.f. = 26$, mean running times 18·2 sec. and

TABLE I

PERCENTAGES OF ALTERNATION FOR REWARDED AND NON-REWARDED RATS FOR EACH BLOCK OF 6 DAYS OF TESTING

	Block				
	1	2	3	4	5
Non-rewarded rats	91·9	92·0	93·0	90·8	95·3
Rewarded rats	94·0	95·2	92·6	94·0	90·5

6·1 sec.). Although the variability of the two groups decreased with repeated testing, the running times of the two groups overlapped on every test day. There was no trend, after the first 3 days of testing, for the running time data over the days of testing.

Partial entries

Although the non-rewarded rats made more partial entries than the rewarded rats (means of 2·7 on trial 1 and 0·4 on trial 2 of the last day of testing compared to 0 on both trials for the rewarded rats), the hypothesis that partial entries accounted for Still's results could not be tested since no difference between rewarded and non-rewarded rats on the spontaneous alternation measure was obtained.

DISCUSSION

This experiment failed to replicate the findings of Still (1966). Normally failure to replicate involves failure to reject the null hypothesis and such an eventuality is not convincing. However, in the present experiment the obtained difference at the end of testing was temporarily in a direction opposite to that predicted by Still and this was at the 10 per cent. level of significance. The present author would place little confidence in this result, but it does convincingly show a failure to replicate Still's result.

It was suggested that Still's finding could have been an artefact of the longer decision times of the non-rewarded rats. Over the last 15 days of testing the mean running time of the non-rewarded rats ranged from 11·7 sec. to 36·0 sec. The range for the rewarded rats was 1·2 sec. to 5·0 sec. The intervals involved here would be unlikely to reduce spontaneous alternation behaviour by virtue of the memory trace decay (Dember and Fowler, 1958). Although the groups did differ in their partial entry behaviour this appeared to have no effect on the alternation behaviour.

The present experiment differed from Still's in a few aspects. The present experiment used a different reward and gradually imposed the final testing conditions. This was done since Frosty-Os are a more palatable food for rats and because it was considered unlikely that a large proportion of the rewarded rats would eat in 15 sec. on test trials early in training. (Informal observation bore this suspicion out.)

The results of the present experiment, therefore, throw doubt on the generality of Still's findings.

DAVID LESTER.

REFERENCES

DEMBER, W. N., and FOWLER, H. F. (1958). Spontaneous alternation behaviour. *Psychol. Bull.*, **55**, 412–28.

FREEDMAN, P. E. (1965). Habituation of alternation behaviour. *J. exp. Psychol.*, **69**, 613–7.

SPEAR, N. E., HILL, W. F., and CAMPBELL, D. J. (1964). Effect of unconsumed reward on subsequent alternation of choice. *Psychol. Rep.*, **15**, 407–11.

SPEAR, N. E., HILL, W. F., and O'SULLIVAN, D. J. (1963). Spontaneous alternation with and without reward in hooded and albino rats. *Psychol. Rep.*, **12**, 441–2.

STILL, A. W. (1966). Spontaneous alternation and pattern of reinforcement. *Quart. J. exp. Psychol.*, **18**, 109–18.

Manuscript received 17th April, 1967.

Lester's conclusions concerning the generality of my results (Still, 1966, experiment 3; the experiments in that paper will be referred to here as E1, E2, and E3) seem well founded. Unrewarded rats do *not*, as I had concluded, usually show a decline in spontaneous alternation with repeated testing. In addition to Lester's result, there is the following experiment (E4) performed in this laboratory:

Fourteen female hooded rats were divided into three groups, and tested in a wooden T-maze identical to that used by Douglas (1966). One group (five rats) was rewarded with food in foodcups at the end of both arms, another group (5 rats) was unrewarded, but found empty food cups smelling of food (having been filled with rat pellets and then emptied), and a third group (4 rats) was unrewarded, and found no food cups. All rats had previously been given six pairs of trials in the same maze without reward, and they were given a further 28 pairs of trials using a procedure similar to that used in E3. Results are shown in Table I, where it will be seen that only the rewarded group shows any sign of a decline in alternation with testing, and that is not significant.

TABLE I

MEAN PERCENTAGES OF ALTERNATION OVER EACH SUCCESSIVE BLOCK OF SEVEN PAIRS OF TRIALS BY REWARDED (R), UNREWARDED WITH FOOD CUPS PRESENT (NR/FC), AND UNREWARDED RATS WITH FOOD CUPS ABSENT (NR)

Group	*No. of subjects*	*Stage of training*			
		1	2	3	4
R	5	97	94	97	83
NR/FC ..	5	97	91	94	97
NR	4	89	96	82	96

In spite of these contrary results, it is not likely, though of course possible, that the results of E3 are due to chance. The individual results of group 0 per cent are shown in Table II. If the correlation for each rat between stages of training and alternation is taken (using Kendall's tau, with correction for ties, Siegel, 1956), it is significantly below chance at the one tailed 5 per cent level in rats 1 and 3, and at the one tailed 1 per cent. level in rats 2 and 4.

TABLE II

INDIVIDUAL FREQUENCIES OF ALTERNATION OVER EACH SUCCESSIVE BLOCK OF FOUR PAIRS OF TRIALS BY THE UNREINFORCED GROUP IN STILL (1966), EXPERIMENT 3.

Rat	*Stage of training*							
	1	2	3	4	5	6	7	8
1	4	4	4	4	4	3	4	2
2	4	4	3	3	4	3	2	2
3	4	3	4	4	3	3	3	3
4	4	4	4	4	4	3	1	3
Totals	16	15	15	15	15	12	10	10

Thus it seems likely that some alternation reducing variable is operating in E3, but not in E4 or Lester's experiment. Since understanding of the action of any such variable will increase our knowledge of spontaneous alternation itself, it is worthwhile considering some of the dependent and independent variables which have been found to correlate with alternation, to see which of them could account for the results of E3. Particular attention will be paid to choice time, since this is something in which rewarded and unrewarded rats are likely to differ.

1. In E1, 2, 3 and 4 there is no significant or systematic correlation between amount of alternation and the median choice time of the second trial of each pair of trials. A significant association in E3 and E4 does emerge, however, if the choice time of each trial upon which a rat fails to alternate (repetition choice times) is compared with the alternation choice times of that rat at approximately the same stage of training. In E4 the median choice time is found of the two alternation choices before and after each repetition choice; this median is then compared with the repetition choice time, and a plus scored if the repetition choice time is greater, a minus if it is less. In E3 the median of the repetition choice times is found for each rat over each block of four pairs of trials where repetitions occur, and is compared with the median of the alternation choice times from the same block; a plus is scored if the repetition median is greater, a minus if it is less than the alternation median. In both experiments, a rat is given an overall plus if he has more pluses than minuses, an overall minus if he has less. In E4 there were nine overall pluses, one minus; in E3 there were ten pluses, two minuses. Combining these results, the number of pluses was significantly greater than the number of minuses by the binomial test ($p < 0{\cdot}001$). In all groups there were more overall pluses than minuses, and there were no significant differences between groups.

2. Alternation in a Y-maze is sometimes reduced if the door is closed behind the rat following choice on the first trial; this only occurs if the delay between trials is very brief (Sheldon, M. H., personal communication). In E3 and E4, unlike E1, the arm door was not closed behind the rat following a choice on the first trial of a pair, a detail of the procedure that was unfortunately omitted from the description given of E3 in Still (1966). In Lester's experiment, the door was closed behind the rat following a choice, and yet a very high rate of alternation was maintained, so that this does not seem a directly relevant variable here.

3. If a rat is given a shock in the stem on the second trial of a pair (Thompson and Higgins, 1958), or pushed to the choice point (Still, 1966*a*) it will tend to repeat its first choice rather than alternate. It has also been found (Still, 1965) that giving an electric shock outside the maze, during a short delay between trials, significantly reduces alternation. Since it also increases choice time, a significant negative association between choice time and alternation is found if the results from the shocked rats and an unshocked control group are combined.

These experiments under 3 are mentioned, not because they have any *direct* relevance to E3, E4, and Lester's experiment, but because they throw light upon the results described under 1 and 2. Giving a shock in the stem before the second trial, pushing the rat in the stem, and giving a shock outside the maze between trials are all operations that are likely to increase the rat's fear or emotionality at the moment of choice. Closing a guillotine door behind a rat, particularly a naïve rat (as in Sheldon's experiment) is also likely to increase fear, so that these results can be explained if it is assumed, as suggested in Still (1966*a*), that a rat will only alternate spontaneously if its emotional state is below a critical level of fear. Since fear is also likely to increase choice time, this principle could explain the association between alternation and choice time described under 1, if there are uncontrolled factors occasionally producing fear at the moment of choice in E3 and E4.

4. Both O'Connell (1964) and Walker (1956) report a *positive* correlation between alternation and choice time in unrewarded rats given free choices on both trials of a pair. If this means that, other things being equal, long choice times favour high alternation, one would expect more alternation by unrewarded rats. It is also possible, however, that the giving of free trials is the crucial factor in producing this positive correlation. Suppose that rats have a more or less stable turn or position preference, and that this preference determines choice time more than any temporary preference induced by the alternation tendency. This means that choice time will be longer when the rat chooses against the stable preference, than when it chooses according to it, irrespective of whether it is alternating or repeating. Now if all trials are free, the first choice of a pair is likely to be in the direction of the stable preference; in order to alternate, the rat must therefore choose against this preference, and hence more slowly than if it repeats. In this way, the positive correlation found in these experiments could be unconnected with alternation as such, and would not be relevant to E3, E4 or Lester's experiment, where the first trial of a pair was always forced.

5. In one experiment (Still, unpublished), a positive correlation between alternation and weight was found. Nineteen rats were kept in activity wheels and fed 12 gm. of food per day. Mean activity ranged from under 1000 to over 10,000 revolutions per day, and

apparently as a consequence of this differential expenditure of energy, weights at the time of testing ranged from 146 to 224 gm. Two delays between trials of a pair were used, 5 sec. and 1 hr., and each rat was given each delay 16 times. The first trial of a pair was forced, equally often to the two arms. The correlation (Kendall's tau) between choice time and alternation was $+0{\cdot}23$ ($p > 0{\cdot}05$), while that between weight and alternation was $+0{\cdot}48$ (two-tailed $p < 0{\cdot}005$). This correlation with weight has not usually been found, though it is probably related to the lack of variability in the choice behaviour of highly motivated rats found by De Valois (1954). The large range of weights in the present experiment would naturally favour such a correlation, provided that the necessary underlying conditions exist. A possible explanation is that rewarded rats are more likely to acquire a dominant turn or position preference under high motivation; this obviously cannot explain the reduction in alternation by unrewarded rats of E3, but it is a possibility that must be allowed for in experiments on spontaneous alternation and reinforcement.

From this discussion, one possible explanation of the difference between the results of E3 and the other experiments emerges. This is that in E3 some uncontrolled factor, not occurring in E4 or Lester's experiment, was producing fear at the time of choice, was more likely in unrewarded rats and towards the end of the experiment, and was sufficiently frequent to produce marked reduction in alternation. One observation bears upon this point. With training, the unrewarded rats of E3 and E4 tended to try to push their noses between the roof and the walls of the maze, while waiting after having made a choice (the rewarded rats, of course, were eating during this time). In E3 the roofs were of Perspex, not very firmly secured and relatively heavy, so that they were noisy and there was a danger of the rat's nose being hurt. This could produce fear, and hence reduced alternation. In E4 and Lester's experiment, the roofs were of wire netting, and were unlikely to cause pain or come down with a bang. Thus the solid Perspex roof could be the crucial independent variable that reduced spontaneous alternation, via the independent variable "fear", among the unreinforced rats of E3.

I should like to thank the Medical Research Council for financial support; the University of Durham Research Fund Committee for a grant to cover typing expenses; and Mr. A. J. Watson, who supervised my Ph.D. thesis, for which some of the experiments mentioned in this paper were performed.

A. W. STILL

REFERENCES

DE VALOIS, R. C. (1954). The relation of different levels and kinds of motivation to variability of behaviour. *J. exp. Psychol.*, **47**, 392–8.

DOUGLAS, R. J. (1966). Cues for spontaneous alternation. *J. comp. physiol. Psychol.*, **62**, 171–83.

O'CONNELL, R. H. (1964). Comparison of alternation and response to stimulus change. *J. comp. physiol. Psychol.*, **57**, 362–6.

SIEGEL, S. (1956). *Non-parametric Statistics for the Behavioural Sciences*. New York: McGraw-Hill.

STILL, A. W. (1965). *Studies of Memory and Spontaneous Alternation in the Rat*. Ph.D. thesis. University of Cambridge.

STILL, A. W. (1966). Spontaneous alternation and pattern of reinforcement. *Quart. J. exp. Psychol.*, **18**, 109–18.

STILL, A. W. (1966*a*). Coercion and spontaneous repetition in rats. *Anim. Behav.*, **14**, 163–5.

THOMPSON, W. R., and HIGGINS, W. H. (1958). Emotion and organised behaviour: Experimental data bearing on the Leeper-Young controversy. *Canad. J. Psychol.*, **12**, 61–7.

WALKER, E. L. (1956). The duration and course of the reaction decrement and the influence of reward. *J. comp. physiol. Psychol.*, **47**, 167–76.

Manuscript received 16*th November*, 1967.

BOOK REVIEWS

Early Behavior: Comparative and Developmental Approaches. Edited by H. W. Stevenson, E. H. Hess, and H. L. Rheingold. London and New York: Wiley. 1967. Pp. xii + 303 75s.

Hess tackles a very sensitive issue in his introduction to this book when he criticizes the average American psychologist for his "fanatic insistence" that the concept of inherited behaviour is useless. This insistence is backed up, as everyone knows, by an argument guaranteed to stop any further discussion of the problem: that it is not pieces of behaviour that are inherited, but genes. We cannot disagree with such a truism, but it is obvious just the same (as Hess seems to be saying) that the argument is a bit of a confidence trick. Objections to "Instincts" have come less from those with delicate linguistic sensibilities as from those who simply don't want to believe in them. J. B. Watson was quite honest about it: "Our conclusion, then, is that we have no real evidence of the inheritance of traits—Let us, then, forever lay the ghost of inheritance of aptitudes, of "mental" characteristics, of special abilities." Faced with this redoubtable conclusion, McDougall became very angry: "It is disconcerting and portentious that a man whose reasoning processes are habitually of the kind illustrated . . . should be widely acclaimed in this country as a leader in a great field of science."

It took a long time for the controversy to get on to a sensible level. The essays in this book, the result of conferences held in 1963 and 1965, give an excellent idea of the methods that are now used to study the development of behaviour in young animals. One point is very clear. If we are going to speculate about the influence upon adult behaviour of learning in the young animal, it would be useful to have information about the learning abilities that young animals actually possess. Three of the eleven chapters in the book are concerned with this problem. Learning in the human infant is discussed in the contributions of Lipsitt and Papoušek, and Campbell writes a challenging essay on avoidance and escape learning in the rat. The main point raised by Campbell is that the difficulty of equating such variables as level of motivation between animals of different ages makes it almost impossible to say anything about their relative learning abilities. The idea that the young learn more quickly than the old is part of modern mythology, but the evidence for it "is grossly deficient, both in scope and excellence." Campbell himself fails to find any difference in the speed of learning between young and adult rats in tasks where they appear similarly motivated.

The chapter by Eibl-Eibesfeldt gives examples of the research methods that can be used to evaluate the different contributions of phylogeny and ontogeny to the development of behaviour. The experimental manipulation of early behaviour is the subject of several other chapters, notably "The development of handedness in cats and rhesus monkeys," by J. M. Warren, J. M. Abplanalp, and H. B. Warren. The authors investigated the effect of giving extensive practice in manipulation to kittens, hoping to find that it played a part in the development of paw preferences. It appears not to have done so; and in fact there is a trend towards *greater* equivalence between paws in cats that have been given the practice early in life.

Usefully reviewed in the remaining chapters are some more familiar topics, for example: "Visual perception in the human infant" (Fantz); "Social responsiveness in young chimpanzees" (Mason); and "Factors influencing the choice of mate in altricial birds" (Klinghammer). The main conclusions of this interesting collection of papers are summarized in the concluding chapter by Harriet Rheingold.

M. J. MORGAN.

Current Research in Motivation. Edited by Ralph Norman Haber. London and New York: Holt, Rinehart and Winston. 1966. Pp. xii + 800. £5.

If you don't like weight on your stomach, this book is not for you. It contains seventy-six papers on motivation, roughly half of which are on animal work and the remainder are on complex aspects of human motivation. Almost all the articles have been previously published and are readily available elsewhere: several are extracts from other books. Although papers on man and on animals are interleaved, the gulf between them is very wide. It is hard to believe that many readers will be interested in both aspects of motivation and it seems a pity that the editor did not concentrate on bringing out a more comprehensive collection of papers on one aspect or the other.

The level of specificity of the articles varies enormously—some articles contain reviews of a field (e.g. Zeigler on Displacement Activities), others report highly specific experimental results unrelated to theory. It is difficult to see at whom the book is aimed since the latter type of article is only of interest to experts and there are not a sufficient number of articles on any one topic to make the book of much value to anyone interested in a particular area of motivation.

The principles of selection are by no means clear. The only significant trend is that there are more articles by members of Yale University than would be expected from a random sampling of the literature ($p < 0{\cdot}001$). A reviewer can always think of articles he would have liked to see included in a collection such as this. Some of the omissions which surprised me are as follows: although there are five articles on instinct, none of them is by an ethologist; very little attention is given to the physiological mechanisms underlying "biological" drives although our understanding of these mechanisms has increased dramatically over the last twenty years; there is almost no mention of self-stimulation; there are no articles by Spence, Stellar or Teitelbaum.

The diversity of the articles both in content and in level of presentation makes it almost impossible for a reviewer to make any general comment. My own impression is that we have made very little advance in conceptual clarity over the last twenty years in this area of psychology. Motivation cannot be divorced from theories about what it is that animals learn so that the attempt to treat motivation in isolation from any clear theory of learning seems to be fruitless. Some important new theoretical concepts have emerged recently, e.g. the role of frustration in animal learning and the effects of cognitive dissonance on human attitudes and behaviour: the attempt to work out the detailed influences of such variables on behavioural phenomena is interesting and fruitful.

Many of the articles on human motivation seem to the present reviewer curiously turgid and uninspired. For example a group of subjects high in *n* Affiliation showed more anxiety when they knew they were being rated for "popularity, attractiveness etc." than a group low in *n* Affiliation. This seems to prove that if you are anxious to be liked, you will be anxious to be liked. The practice of inducing in male subjects high arousal of sexual feeling ("HAS") by showing photographs of naked girls ("nude females") is increasing, but it is reassuring to discover that photographs of female models only induce a low arousal of sexual feeling ("LAS").

This book can be strongly recommended to anyone who has £5 to spare and who happens to want a random selection of articles on motivation.

N. S. SUTHERLAND.

Language and Thought. Edited by Donald C. Hildum. London: Van Nostrand, Insight Book No. 37. 1967. Pp. vi + 200. $1.95. 16s.

For whom are books of readings intended? From their content it can usually be deduced that it is neither the introductory course student nor the graduate student/specialist. It may be the student in higher undergraduate courses, or the specialist in another area who wants to keep in touch and buys a book knowing he will feel obliged to read it.

Hildum's book will serve the undergraduate taking a course in psycholinguistics moderately well, although he will probably also read many of the sources. The book is an introduction to linguistics for psychologists, in effect. The "Thought" in the title does not refer to non-linguistic cognition, but to the inclusion of semantics. Although a few germane excerpts are included, in places the selection and sequencing is poor and the brief, non-interpretative comments do not improve matters. Why, for example, begin the book with your old pal the Magical Number Seven, rather than G. A. Miller's paper on "The Psycholinguists" from "Encounter"; or if "cognitive economy" must be shown, why not Garner's elegant justification of the binary principle? There should either have been much more editorial cutting and comment, or none at all, depending upon whether a book with a theme or an overview was intended. But the Series Editors, not Hildum, are to blame for this.

MARK P. HAGGARD.

Hormones and Behavior. Edited by Richard E. Whelan. London: Van Nostrand. Insight Book No. 39. 1967. Pp. vi + 266. $2.95. 24s.

This book is a collection of eighteen articles, presented in four groups. The first deals with effects of stimuli from individuals of the same species, such as a dominant mouse, a castrated male dove, and newborn rat pups, on hormonal secretion. The second group relates behaviour patterns, such as fish migration and the sexual behaviour of rats, cats, capons, and man, to the presence of particular hormones, and where possible to the amount of hormone necessary for the occurrence of the behaviour and to the mechanisms underlying hormonal action. The last two groups investigate how hormones affect either later or learned behaviour. Three papers show that gonadal or thyroid hormones in the fetus or neonate determine behaviour at a later stage: androgen administered prenatally to female guinea pigs determines later responsiveness to exogenous hormones; lack of androgen in male neonatal rats later results in feminization; and impairment in a closed field test is greatest in rats thyroidectomized at birth rather than later. The last four papers show that thyroid and adrenal hormones are secreted during conditioned avoidance learning, and do affect performance, although how this occurs is still not clear.

It is unfortunate that only one paper deals with the integration of behaviour over a long period of time, so that interactions between hormonal condition, other physiological changes, behavioural changes, and external stimuli are neglected. However, Whalen has sampled a wide range of species, hormones, and behaviour, and has arranged the material into meaningful groups. This, together with a brief, but not over-simplified, summary of endocrine and endocrine/behaviour relationships, should make this a helpful book for students of behaviour.

JOAN G. STEVENSON.

PUBLICATIONS RECEIVED

BOOKS

Science and E.S.P. Edited by J. R. Smythies. London: Routledge & Kegan Paul. 1967. Pp. viii + 306. 42s.

Second Thoughts: Selected Papers on Psycho-Analysis. By W. R. Brion. London: Heinemann Medical Books. 1967. Pp. iii + 173. 25s.

Problems of Psychotherapy. By Herbert Zucker. London: Collier-Macmillan. 1967. Pp. viii + 194. 48s.

Interpersonal Styles and Group Development: An Analysis of the Member-Leader Relationship. By Richard D. Mann in collaboration with Graham S. Gibbard and John J. Hartman. London and New York: Wiley. 1967. Pp. x + 308. 62s.

William James: A Biography. By Gay Wilson Allen. London: Rupert Hart-Davis. 1967. Pp. xx + 556. 84s.

An Atlas of Figure Drawings: Studies on the Psychological Characteristics of Medical Students. III. By Caroline Bedell Thomas. London: Oxford University Press. (John Hopkins Press) 1967. Pp. xvi + 922. £10.

Chemotherapy of Psychiatric Disorders. By Peter Dally. With an Introduction by William Sargant. London: Elek Books. 1967. Pp. viii + 156. 42s.

PAPERBACKS

Psychologie Sociale Experimentale. By R. B. Zajonc. Paris: Dunod. 1967. Pp. x + 139. 18F.

Morale of the Dublin Busmen: A Sociodiagnostic Study of the Dublin City Services of Coras Iompair Eireann. By Hans van Beinum. Dublin: Irish National Productivity Committee. 1967. Pp. 94.

Automation: Its Anatomy and Physiology. By John Rose. London and Edinburgh: Oliver & Boyd. Contemporary Science Paperbacks. 1967. Pp. vi + 144. 7s. 6d.

Automation: Its Uses and Consequences. By John Rose. London and Edinburgh: Oliver & Boyd. Contemporary Science Paperbacks. 1967. Pp. vi + 123. 7s. 6d.

Psychology and Social Problems. By Michael Argyle. London: Tavistock Publications; New York; Barnes & Noble. Social Science Paperbacks. 1967. Pp. 232. 16s. Bound Edition—Methuen. 25s.

PROCEEDINGS OF THE EXPERIMENTAL PSYCHOLOGY SOCIETY, 1967

29th–31st March, 1967. Meeting at Keele.

1st Session: "Size and distance judgements under water and in fog," by Dr. Helen E. Ross. "The role of vision in static and mobile balancing," by Mr. J. Dickinson.* "The green and red sensitive mechanisms in dark adaptation," by Mr. W. Ernst.* "The effect of self-instruction on performance," by Dr. Sheila Jones.*

2nd Session: "Visual perception and evoked brain potentials," by Dr. D. Regan.* "Human occipital and retinal potentials evoked by subjectively faded visual stimuli," by Mr. L. A. Riggs* and Dr. P. Whittle. "Psychophysical indications of orientation-detecting units," by Dr. D. P. Andrews.*

3rd Session: "Pupil size as a measure of information processing in visual tasks," by Mr. J. L. Bradshaw.* "Multi-channel information processing" by Mr. E. Whelan.* "Yet more experiments in two-channel listening," by Dr. N. Moray.

4th Session: "Subjective simultaneity and the perceptual moment," by Mr. D. A. Allport.* "Decision times in signal detection," by Dr. G. Wolfendale.*

5th Session: "Computer analysis of inter-response times in operant conditioning schedules," by Mr. R. T. Bradley.* "The role of intelligence in the perception of causality," by Mr. N. A. Beasley.*

10th–12th July, 1967. Meeting at Oxford.

1st Session: "Theory of word selection," by Dr. Davis Howes. "The apparent reversal of rotary motion in depth," by Dr. R. P. Power.* "Delayed paired comparison of verbal or non-verbal stimuli presented tachistoscopically to the right or left visual hemi-field," by Dr. A. W. H. Buffery.

2nd Session: "E.C.S. and retrograde amnesia," by Dr. Sheila Zinkin.* "Organization and memory," by Dr. G. Mandler.* "What have we learned from probability-learning?" by Dr. M. Friedman.*

3rd Session: "The relation between reaction time and error at different levels of confidence," by Mr. J. Wilding.* "The refractory period of C-reactions," by Dr. P. Bertleson and Mlle. Francoise Tisseyre.* "Is there a motor constancy mechanism?" by Dr. D. Legge and Dr. Judith Pottinger.* "Changes in auditory threshold due to an auditory warning stimulus," by Dr. J. Armstrong.*

4th Session: "Mach bands and the mechanism of colour vision," by Mr. M. L. Mathews.* "Some factors affecting the action of mescaline on operant behaviour," by Dr. J. R. Smythies, Mr. V. S. Johnson* and Mr. R. T. Bradley.* "The effect of brain and spinal lesions on prism adaptation in monkeys," by Dr. J. Bossom.* "Electro-physiological and behavioural studies of the superior colliculus in mammals: what the frog's eye tells the monkey's brain," by Dr. N. K. Humphrey* and Dr. L. Weiskrantz.

4th–5th January, 1968. Annual General Meeting at Birkbeck College, London.

1st Session: "Reshaping visual co-ordination," by Dr. I. P. Howard. "Adaptation of divers to optical distortion," by Dr. Helen E. Ross. "Visual delay as a function of luminance," by Mr. J. A. Wilson. "Osgood's congruity principle and semantic differential judgement of colour," by Dr. J. H. Hogg.*

2nd Session: "A comparison of the effects on depression and memory of bilateral E.C.T. and unilateral E.C.T. to the dominant and non-dominant hemispheres," by Dr. A. M. Halliday, Dr. K. Davidson,* Dr. M. W. Browne* and Dr. L. C. Kreeger.* "The complication experiment simplified," by Dr. P. Bertelson.

3rd Session: "Refractory period piece," by Dr. L. H. Shaffer.* "Eye and ear as independent observers," by Dr. D. W. J. Corcoran. "Doubts about the Sperling store," by Dr. D. H. Holding.

4th Session: "An auditory negative after-image," by Dr. J. P. Wilson.* "The after-effect of errors in serial choice response tasks," by Dr. P. M. A. Rabbitt. "Subjective probability in choice-reaction experiments," by Dr. D. R. J. Laming.

* By invitation.

Committee, 1968: President: Professor J. Brown.

Editor: Mr. A. J. Watson.

Ordinary members of the Committee: Dr. J. Annett, Dr. F. I. M. Craik, Dr. R. Davis (Hon. Treasurer), Dr. B. M. Hermelin, Dr. D. Legge (Hon. Secretary), Dr. P. M. A. Rabbitt (Asst. Hon. Secretary), Dr. I. S. Russell, Dr. A. W. Still, Dr. R. T. Wilkinson.

The following have been elected into membership of the Society:

D. E. Blackman, Ph.D., Department of Psychology, The University, Nottingham.
T. M. Bloomfield, Ph.D., Laboratory of Experimental Psychology, University of Sussex, Falmer, Brighton.
D. A. Booth, Ph.D., Laboratory of Experimental Psychology, University of Sussex, Falmer, Brighton.
M. P. Haggard, Ph.D., Psychological Laboratory, Downing Street, Cambridge.
G. J. Harrison, Ph.D., Department of Psychology, The University, Sheffield S10 2TN.
P. N. Johnson-Laird, Ph.D., Department of Psychology, University College London, Gower Street, London, W.C.1.
D. J. MacFarland, D.Phil., Institute of Experimental Psychology, 1 South Parks Road, Oxford.
E. M. McPhail, D.Phil., Laboratory of Experimental Psychology, University of Sussex, Falmer, Brighton.
J. C. Marshall, Ph.D., MRC Speech and Communication Research Unit, The University, 31 Buccleuch Place, Edinburgh 8.
L. H. Shaffer, Ph.D., Psychological Laboratory, University of Exeter, Gandy Street, Exeter.
W. B. Templeton, M.Sc., Department of Psychology, The University, 7 Kepier Terrace, Gilesgate, Durham.
J. P. Wilson, Ph.D., Department of Communication, The University, Keele, Staffs.

Visiting Foreign Member:

Dr. J. Dewson, Institute of Experimental Psychology, 1 South Parks Road, Oxford.

QUARTERLY JOURNAL
OF
EXPERIMENTAL PSYCHOLOGY

ANNOUNCEMENT

Publication Arrangements

With effect from 1st January, 1969 the Quarterly Journal of Experimental Psychology will be published for the Experimental Psychology Society by Academic Press Inc. (London) Limited. Subscribers will receive renewal notices from the Academic Press. All transactions to take effect after 1st January, 1969 should be conducted with:

Academic Press Inc. (London) Limited,
Berkeley Square House,
Berkeley Square,
LONDON, W.1

or through booksellers as hitherto.

THE QUARTERLY JOURNAL OF
EXPERIMENTAL PSYCHOLOGY

Vol. XX 1968 Part II

LIBRARY

THE PERCEPTION AND LEARNING OF TEMPORAL PATTERNS

BY

W. R. GARNER and RICHARD L. GOTTWALD

From Departments of Psychology, Yale University and Johns Hopkins University

Subjects observed sequences of stimuli generated by the repetition of temporal patterns formed of eight elements, each element being a binary event. Rate of presentation varied from 0·8 to 8·0 elements per sec.; stimuli were either visual or auditory; there were 10 different basic patterns; and for each basic pattern there were a preferred and a non-preferred starting pattern. Subjects observed as long as necessary in order to give a verbal description of the pattern after terminating it. Difficulty (measured by observation time and accuracy of description) was related to basic pattern and generally increased with increased rate. Differences due to starting pattern were found only at low rates, while differences due to modality were found only at high rates. Similar effects were obtained with respect to pattern description.

These consistently different effects of starting pattern and modality at low and high rates require a distinction between *pattern perception* at higher rates, which is phenomenally integrated, immediate, compelling and passive, and *pattern learning* at lower rates, which is unintegrated, derived, intellectualized and active. Even though a distinction between perception and learning is necessary, similar principles of pattern organization operate for both.

INTRODUCTION

The simplest way to generate a temporal pattern is to produce a series of binary events of fixed length. If such a pattern is continued without interruption into a continuous sequence, then we have to distinguish between two aspects of the pattern: the *basic pattern* and the *specific pattern*. A specific pattern is any particular combination of binary events of the given length, e.g. the pattern 00110001 for a length of eight elements. Another specific pattern might be 11000100. Each of these two patterns, however, when continued indefinitely and without interruption is the same, since both patterns produce the same continued sequence of elements, and differ only in the nature of the pattern as started. Thus we can consider them to be two representations of the same basic pattern.

In general there will be as many specific patterns for each basic pattern as there are elements in the pattern, except for a few cases in which the basic pattern degenerates to one of shorter length, e.g. 00110011, which degenerates to a length of four rather than eight.

This difference between basic and specific patterns can be used in the psychological study of temporal pattern in two ways. First, we can use the specific pattern as an experimental variable by starting sequences with different specific starting patterns, to determine the effect of such differences on the nature of the perceived pattern and on the difficulty of perceiving or learning a particular pattern. Second, we can use the specific pattern as a dependent variable by determining which specific pattern is learned or perceived for a given basic pattern. In such a case our interest is not only in the actual pattern perceived, but also in the number of different specific patterns which will be perceived or learned for a given basic pattern.

Garner (1962, pp. 202–205) first suggested that the number of alternative specific patterns is related to the perceived goodness (in the Gestalt sense) of these patterns, and Garner and Clement (1963) showed that the judged goodness of spatial dot patterns was greater when there were (both subjectively and objectively) fewer alternative patterns for the pattern being judged. In other words, good patterns were shown to have few alternative organizations. Extensions of this idea were made by Clement (1964), in showing that the number and latency of verbal associates to dot patterns decreased with better patterns: and by Handel and Garner (1966), in showing that the number of other dot patterns which would be used as associates to a given pattern was smaller with better patterns.

In a study of temporal patterns, Royer and Garner (1966) used sequences of dichotomous sounds produced by repeating patterns of eight elements at a rate of two elements per sec. Their subjects listened to a sequence until they were able to track it (using two mechanical keys), and they were allowed to begin tracking the sequence at any point in the pattern. The point at which responding began was generally at the beginning of a long run or a series of single alternations. Moreover, the variability of the point of responding (thus the variability of the subjective pattern organization for any basic pattern) increased with pattern complexity, as did also task difficulty as measured by delay before responding and errors made after responding began. In addition, these authors used all specific patterns as starting patterns for each basic pattern and found that starting a sequence with a subjectively nonpreferred pattern made the task more difficult, although it had little effect on the point at which responding began. Thus using different starting patterns affected the difficulty of the task but not the subjective organization.

Garner and Gottwald (1967) used sequences of two lights, the length of their basic pattern being five elements, and the presentation rate being one element every 4 sec. The subject was required to learn the pattern, begin responding to it in anticipation of each element, and then to describe the pattern after learning criterion had been met. The factors which determined the difficulty and organization of their sequences were similar to those in the Royer and Garner study. However, there were also some differences. Responding began at points in the sequence corresponding to the preferred organization (as evidenced by descriptions) only if the first element in the pattern had a high probability of being correct. Furthermore, the starting pattern had more influence on subjective organization, so that if a sequence was started with a nonpreferred pattern, that pattern would sometimes be used as the description; while if the starting pattern was a preferred pattern, then it was almost invariably used as the description. Still further, a nonpreferred starting pattern produced learning delay similar to that found by Royer and Garner, and the delay was to some extent due to the fact that the particular starting pattern induced expectancies about the basic pattern which were incorrect in some cases and thus led to high error rates.

Ordinarily we speak of "learning" with stimuli presented at rates such as one element every four seconds, although the term "perception" seems more appropriate for the faster rate used by Royer and Garner. Certainly, the subjective experience with these faster rates is one not of learning but of perceptual organization. In fact, once a sequence becomes organized into successive patterns with clear beginnings and endings, the perceptual experience is so definite that one is left with a sense of incompletion if the sequence is not terminated at its natural ending. Further evidence of the importance of rate with respect to pattern perception has been provided by Royer (1967) in showing that the maximum key-tapping rate which can be produced depends both on the basic pattern and the specific pattern within the basic pattern.

Thus while these experiments indicate that there is much in common between slow "learning" and fast "perception," the two processes are far from identical, and further exploration of these processes as a function of rate of stimulus presentation is required.

The experiment of Royer and Garner differed from that of Garner and Gottwald not only in the rate of stimulus presentation, but also in that different modalities were used: auditory in the former experiment and visual in the latter. If the phenomenon being studied is truly one of pattern perception, rather than auditory or visual perception, then we have no reason to expect this factor to have had any effect on the different results obtained. Klemmer (1967), however, has shown that the precision with which sequences of responses can be produced to either fixed- or variable-rate stimuli is better with auditory than with visual stimuli. Thus it is possible that the modality used may have some effect on performance related to either pattern perception or learning.

Purpose of experiment. For these various reasons the purpose of the present experiment is to investigate temporal pattern experience as a function of the basic and specific pattern, at various rates of stimulus presentation, and with both visual and auditory stimuli.

METHOD

Apparatus

The apparatus consisted of a vertical board on which were mounted two lights, 11 in. above a table which was itself 34 in. high. Each light consisted of a coloured glass transilluminated by an incandescent bulb. Each coloured glass light was 1·0 in. in diameter, with a brightness of approximately 0·3 foot lamberts. They were placed side by side, 2·0 in. apart at the centres, and the light to the left of the subject (as he faced the display about 2 ft. from it) was green, and that to the right was red. Thus the binary light events were distinguishable by both colour and location.

Two doorbell buzzers were mounted on the back of the board, 18 in. apart. These buzzers had an intensity of 75 dB. sound pressure level. The fundamental frequency of the buzzer to the left of the subject was 145 cps, and that to the right was 195 cps. The two buzzers also differed considerably in harmonic structure, and for many subjects this qualitative difference was more obvious than the pitch difference. Thus the binary sound events differed in pitch, quality and location.

The light or sound stimuli could be cut off by the subject at any time by his pressing a telegraph key located in front of the display, and if necessary could be restarted by pressing another key. The stimulus sequences were generated with a set of cams which the experimenter could select, and these cams were cut so that the duration of each stimulus event was equal to the duration of the blank period between events. A synchronous motor drove the cams, and rates were varied with mechanical gears. Selection of modality was accomplished with electrical switches. The electrical output controlling the stimulus presentation was monitored on a continuous event recorder so that length of stimulus presentation could be determined.

One subject at a time and the display equipment were located in a soundproof room to eliminate the undesired sound of the control apparatus outside the room, and the subject and experimenter communicated through an intercom system.

Task

The subject's task was very simple. When a particular stimulus sequence was started (always either sound or light, never both), the subject observed until he thought he could describe the pattern. At that time, he stopped the stimulus presentation and described the pattern verbally to the experimenter, in terms of "left" and "right," a description appropriate to either light or sound stimuli. The experimenter said "thank you," but did not correct any errors. If the subject forgot the pattern while describing it, he was able to start the sequence again. This action rarely occurred. If the subject did not stop the sequence after 640 events had been presented, the experimenter stopped it and asked the subject to describe the pattern. The experimenter had to stop the pattern only at the fastest speed, and in only one such case did the subject then give a correct description. Total observation time per subject was approximately 1 hr.

For the patterns used in this study, Royer (1967) found that the maximum rate of responding ranged from 3·3 to 4·9 elements per sec. Since we used rates nearly twice as fast as these, the tracking method used by Royer and Garner (1966) was impossible.

Experimental variables

Basic patterns. The patterns were all eight events long and consisted of 10 of the basic patterns used by Royer and Garner. They are shown in Tables I and II, with "0" standing for left and "1" standing for right as the patterns were actually presented. None of these 10 basic patterns degenerates to a shorter length, and only one of them consists of just 2 runs. Thus these constitute a good range of fairly complex patterns.

Starting pattern. For each basic pattern, two starting patterns were selected, based on frequency with which the specific pattern was used as a beginning response pattern in the Royer and Garner study. A *preferred* starting pattern was selected which had been most used in the previous study; a *nonpreferred* starting pattern was selected which had been either the least used or at least rarely used in the previous study. The particular preferred and nonpreferred starting patterns for each basic pattern are shown in Table II.

Rate of presentation. Five different presentation rates were used: 0·8, 1·6, 2·67, 4·0, and 8·0 elements per sec. Thus at the fastest rate the entire pattern was presented once a second.

Modality. As previously indicated, the patterns were presented as either *auditory* or *visual* stimuli.

Experimental design

These experimental variables produce a total of 200 different experimental conditions (10 basic patterns, 2 starting patterns, 5 rates, 2 modalities). These 200 conditions were divided into five sets of 40 each and nine different subjects (from introductory psychology course at Johns Hopkins University) were assigned to each of the five sets of conditions.

Each set of 40 conditions was selected to provide an orthogonal arrangement of the 20 conditions defined by 2 modalities, 2 starting patterns and 5 rates. All 10 of the basic patterns were then assigned to each of the four conditions defined by modality and starting pattern, two to each rate; and these assignments were arranged so that no basic pattern was used more than once for each rate. This design means that each subject observed each basic pattern four times, twice with each starting pattern, and twice with each modality.

The five different groups provided a counterbalancing of the assignments of basic patterns to other conditions so that all combinations of all four experimental variables were used, with nine subjects thus providing data for each of the 200 experimental conditions.

Order of presentation of conditions was arranged so that visual and auditory presentations alternated, speeds progressed from fast to slow or vice versa, no pattern was presented within five trials of itself, and each modality and rate was used for the first trial an equal number of times. Two such orders were used for each of the five sets of conditions, five subjects receiving one order and four subjects the other.

While this overall experimental design is quite complicated in order to counterbalance for various factors we felt might produce artefacts in the results, what actually happened for any given subject was not complicated, and a description can serve as a recapitulation of the design. One particular subject was given a sequence of 40 trials in which visual and auditory presentation alternated. A series of five trials progressed regularly either in increasing or decreasing speed; for half the subjects speeds always increased and for the other half they decreased, although all different speeds were used for the very first speed, the progression commencing after that. Then over the 40 trials, all 10 of the basic patterns were used four times each, in random order; and each basic pattern was started twice at its preferred starting position and twice at its nonpreferred starting position, and each of these occurred once with each modality, the arrangements of modality, basic pattern, and starting point being random except for these restrictions.

Thus, except for the very first trial, each subject knew what modality and what speed to expect, so that surprise factors would be minimized with respect to these. Then over the course of the experiment for him, he heard or saw each of 10 basic patterns four times, once with each combination of starting point and modality. This was a sufficiently small number of experiences with each pattern that we felt the task would not reduce to one of recognition rather than initial perception or learning.

Before commencing the experiment proper, the experimenter demonstrated the task to the subject, and showed him several patterns at different speeds and modalities. In addition, the subject was told that lights and buzzers alternated and that rates progressed. Thus the subject was aware not only of the general nature of the patterns, but before each presentation knew the modality and approximate rate.

RESULTS*

The two primary data obtained from each subject on each presentation were the number of elements or events observed and the pattern description. The median number of elements observed by the nine subjects for each condition was used as the basic summary measure, and geometric means of these for a summary measure of various combinations of conditions. These measures were used because of the very skewed distribution of scores both within and between conditions. The pattern descriptions provided a percentage measure of accuracy, as well as distributions of actual descriptions. The length of observation and average accuracy are measures of *pattern difficulty*, and the actual descriptions and uncertainty or variability of descriptions are measures of *pattern organization.*

Pattern difficulty

Rate. The effect of rate on the number of elements observed before an attempted description is shown in Table I and Figure 1. Over the range of presentation rates used in this experiment, there is a monotonic increase in the number of elements observed as the presentation rate is increased. It would seem entirely reasonable, however, that for slower rates than those used here there would be little or no further

TABLE I

NUMBER OF ELEMENTS OBSERVED BEFORE AN ATTEMPTED DESCRIPTION FOR DIFFERENT BASIC PATTERNS AND RATES OF PRESENTATION

Pattern	*Rate (elements per sec.)*					*Geom. Mean*
	0·8	1·6	2·67	4·0	8·0	
11111000	23·7	32·6	39·1	67·3	124·7	48·0
01101111	20·1	26·1	40·8	65·3	255·9	51·4
11101000	20·5	32·1	50·1	86·5	293·8	60·8
01010111	20·6	29·2	44·3	123·3	316·2	63·5
11110100	25·8	36·3	52·5	84·9	275·4	64·9
11101100	24·9	30·7	46·1	117·8	309·0	66·2
00110111	23·7	39·9	57·8	86·7	271·6	66·4
00101111	27·9	40·7	59·7	99·3	214·8	67·9
11011010	23·2	35·9	59·4	143·9	246·0	70·6
01010011	21·1	38·6	68·6	105·4	385·5	74·3
Geom. Mean	23·0	33·9	51·1	95·3	259·4	62·8

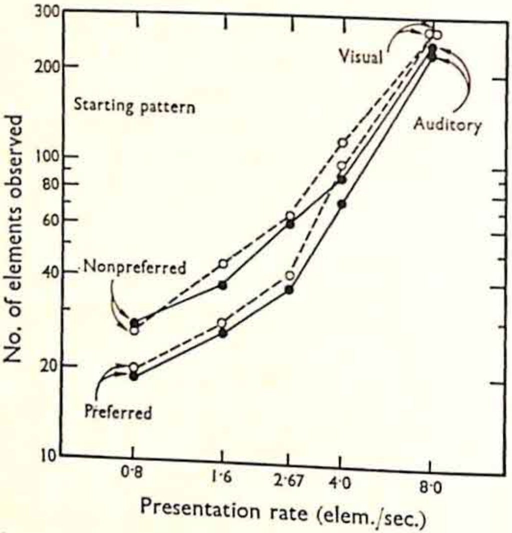
Each cell entry is the geometric mean of medians from both modalities and both starting patterns. Each basic pattern is shown as its preferred starting pattern.

---

* Complete analyses of variance were carried out on all measures by number of elements observed, logarithms of this number, and number of correct descriptions, and statements of effects are justified by statistical significances obtained from these analyses, at least at the 0·05 level.

decrease in the number of elements observed, since the mean for the easier patterns is just about two and one-half complete cycles of the basic pattern, and little further decrease would be possible.

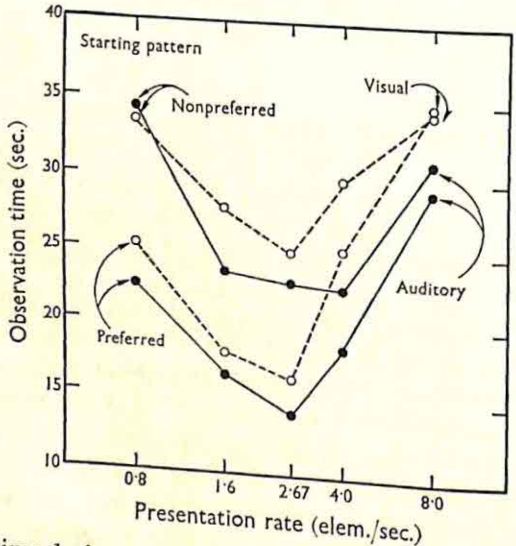
FIGURE 1



Number of pattern elements observed before an attempted description as a function of the rate of presentation of the pattern. (Functions are for both modalities and both starting patterns as indicated. Each plotted point is the geometric mean of medians from the 10 different basic patterns.)

The number of elements observed increases to very high values at the higher rates, being the equivalent, on the average, to 32 complete cycles of the pattern, and being the equivalent of 48 complete cycles for the most difficult pattern. It is clear that rates much higher than the maximum of 8·0 elements per sec. used in this experiment would lead to an almost complete breakdown of pattern perception and the number of elements observed would no longer be a meaningful measure.

FIGURE 2

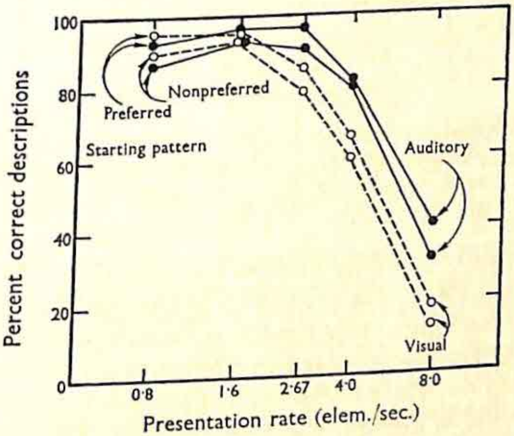


Mean observation time before an attempted description as a function of the rate of presentation of the pattern. (These functions are derived from those of Figure 1.)

Duration of observation in terms of actual elapsed time rather than number of elements is shown in Figure 2, and a somewhat different picture of effect of presentation rate emerges. As presentation rate increases, observation time first decreases and then increases, showing a minimum time at a rate between 2 and 3 elements per sec. Klemmer (1957) has shown that subjects have difficulty maintaining consistent relations between a response and a stimulus when the stimuli are presented at about two per sec. While no continuous overt response was required in the present experiment, it does seem likely that there is a change from the perception of individual elements to the perception of groups of elements at about 2 or 3 elements per sec.

Figure 3 shows the relation between presentation rate of the stimulus elements and the percentage of correct descriptions, and confirms the results found with number of elements observed, namely, that perception is more difficult at higher rates. This measure is somewhat less sensitive than duration because of the effective top value near 100 per cent.; thus there appears to be little effect of rate of presentation at the 2 or 3 lower rates used when accuracy of description is the measure.

FIGURE 3



Percentage of correct descriptions of the ten basic patterns as a function of the rate of presentation of the pattern. (Functions are for both modalities and both starting patterns as indicated. Each plotted point is the arithmetic mean from the 10 different basic patterns.)

Basic pattern. The average number of elements observed varies considerably from one basic pattern to another, as shown in Table I. This result confirms the Royer and Garner (1966) finding of considerable differences in difficulty of perceiving or learning these patterns, and the rank correlation for these 10 basic patterns between number of elements observed in this experiment and response delay in the earlier experiment is 0·63.

There is a substantial interaction between the effects of basic pattern and the effects of rate on number of elements observed, as shown in Table I. The nature of the effect is that differences due to pattern become much greater with the higher rates of presentation. For example, at the lowest rate the average number of elements observed ranges from 20·1 to 27·9, while at the highest rate the average ranges from 124·7 to 385·5. These differences in range are so great that the interaction does not disappear even if ratios rather than differences of values are used.

Table II shows the percentage of errors made in describing the different patterns, averaged over all rates and conditions, and again confirms the result found with number of elements observed. These two measures of pattern difficulty rank correlate 0·63 across patterns. This correlation would have been substantially higher except that pattern 11111000, which required the fewest elements observed, produced a fairly high error rate; and casual observation indicates that this high error rate is largely due to simple carelessness in description, with subjects often describing only four elements of the longer run rather than the required five.

TABLE II

Per cent. Use of Different Possible Descriptions of the 10 Basic Patterns

Description				
Preferred starting pattern	*Second preference*	*Nonpreferred starting pattern*	*Other correct descriptions*	*Errors*
67·2 (11111000)	6·1 (00011111)	0·0 (11100011)	0·0	26·7
38·9 (01101111)	15·0 (11110110)	21·1 (01111011)	5·0	20·0
61·1 (11101000)	5·6 (01000111)	7·2 (10001110)	6·1	20·0
45·0 (01010111)	27·2 (11101010)	2·2 (11010101)	4·4	21·1
55·6 (11110100)	11·1 (01001111)	7·7 (10011110)	2·8	22·8
65·5 (11101100)	6·1 (01100111)	0·6 (10110011)	5·0	22·8
40·0 (00110111)	17·8 (11100110)	1·1 (11001101)	13·9	27·2
35·6 (00101111)	37·2 (11110010)	0·0 (11001011)	5·6	21·7
35·6 (11011010)	4·4 (01101011)	9·4 (01101101)	6·1	44·4
37·8 (01010011)	17·8 (11010100)	3·3 (10011010)	13·9	27·2
48·2	14·8	5·3	6·3	25·4

Numbers show the percentage of each type of description for each basic pattern, the last row being means for all patterns. The actual description is given in parentheses. Errors include failures and incorrect descriptions. Average data for all rates, modalities and starting patterns.

Starting pattern. The important effect of starting pattern is due almost entirely to its interaction with rate of stimulus presentation. Figure 1 shows the effect on number of elements observed. At low rates the use of a nonpreferred starting pattern increases pattern difficulty compared to a preferred starting pattern, while at high rates there is essentially no difference due to starting pattern. Figure 2 shows this same effect when time of observation, rather than number of elements, is the measure. At the lowest rate the average observation time with the nonpreferred starting patterns is approximately 10 sec. greater than with the preferred starting patterns, while there is practically no difference at the highest rate. Figure 3 shows this same effect when per cent. correct descriptions is the measure, although the effect is less clear-cut because of the limiting nature of this measure. Nevertheless, preferred starting patterns give higher average accuracy than nonpreferred starting patterns at the two lowest rates, and this effect is relatively less at the higher rates.

Thus overall these results show that use of a nonpreferred starting pattern at low and even at medium rates produces greater difficulty of learning and perceiving compared to the use of a preferred starting pattern. This result confirms that of Royer and Garner (1966) and Garner and Gottwald (1967). An entirely new result is that there is very little effect of starting pattern on performance at high rates.

Modality. The important effect of modality is likewise due almost entirely to its interaction with rate of stimulus presentation, although the effect is quite different from that found for starting pattern. Figure 1 shows that there is practically no difference in number of elements observed for the two modalities at low rates, but there is a clear effect at high rates. In linear measure, at the highest rate the visual stimulus required 275 elements before a description was attempted, while the auditory stimulus required 245 elements. Figure 2 shows the same effect with observation time as the measure, with little effect of modality at the lowest rate, but a difference of about 4 sec. at the highest rate, the visual stimulus requiring more time. Still further, Figure 3 shows a clear effect of modality on accuracy at the highest rates, but practically no effect at the lowest rates.

Thus the effect of modality on pattern difficulty is in contrast to that of starting pattern, in that starting pattern has its effect only or primarily at low rates, while modality has its effect primarily at high rates.

In effect, this result showing that the interaction of modality with rate is different from the interaction of starting pattern with rate, is a result showing a triple interaction between rate, starting pattern and modality. However, it seems more meaningful to describe the separate two-factor interactions than a single three-factor interaction. The significance of the triple interaction can be demonstrated with a simple nonparametric calculation. At the lowest rates, the four conditions defined by starting pattern and modality form two groups, one for each starting pattern, and our previous research has demonstrated that what differences exist at lower rates should be related to starting pattern, and in the direction found. There are six different ways in which these four conditions could have formed two groups of two each, with one group being higher in performance than the other. Since the particular outcome is just one of six which could have occurred, we can state that its chance probability is 1/6. Klemmer (1967) showed that auditory stimuli provide better performance in a timing task at higher speeds than do visual stimuli, so the result that differences at higher rates are due to modality is predictable. Again the four conditions form two groups of two, in an order expected from previous research, so the probability of this outcome occurring by chance is also 1/6. The joint probability that the expected result should occur at both high and low rates is then 1/36. The systematic change between the effects of modality and starting pattern with increasing rate could not have been known, but the limiting conditions are quite in accord with other known results.

Pattern organization

Pattern description is the primary measure of pattern organization, with variability or uncertainty of pattern description being an important secondary measure. Both Royer and Garner (1966) and Garner and Gottwald (1967) showed that uncertainty of description or response starting point were related to pattern complexity and difficulty, with more uncertainty being associated with greater difficulty. In those experiments, all possible starting patterns had been used. In the present experiment, however, only two starting patterns were used with each basic pattern, and these were selected to show maximum effects of starting pattern. The result is that overall measures of pattern uncertainty are not too useful in the present case, although the nature of the results makes clear that the previously found relations between pattern uncertainty and difficulty also hold in this experiment.

For each basic pattern there are eight different correct descriptions, one starting at each element in the pattern. For purposes of discussion, it is useful to classify these eight possible descriptions into four types, and these four types can then be generalized across basic patterns: (1) the preferred starting pattern; (2) the second preference as obtained from these data; (3) the nonpreferred starting pattern as used in this experiment; and (4) the other five correct responses. The specific patterns for each of the basic patterns corresponding to these classes are indicated in Table II, along with the per cent. usage of each of them.

Rate. There are no important average effects of rate on description. Rather, the effect of rate depends almost entirely on its interactions with starting pattern and modality.

Basic pattern. The per cent. use of each of the four types of pattern description are shown separately for each pattern in Table II. In every case except one (00101111) the preferred starting pattern was also used most frequently as the description. Since the selection of the preferred starting pattern was made from the Royer and Garner (1966) study, these results show that there is considerable generality in pattern organization across different experimental conditions. The second preferred description in the present experiment can also be compared to that found in the Royer and Garner study, and the second preference found in the present study corresponds to those previously found in every case except that the specific pattern 01011011 was preferred in the previous study to the specific pattern 01101011 found as second preference in this study. The nonpreferred starting pattern was used as a description only 5·3 per cent. of the time on the average, and in only one case (01111011) was it used more than 10 per cent. of the time. All other correct descriptions were likewise seldom used.

Overall these results show a very high degree of correspondence with the previous results, and indicate that the basic factors which determine pattern organization are quite general. For these patterns, the two primary factors are the existence of longest runs as either the beginning or the end of the pattern, or the existence of a series of single alternations as the beginning of a pattern, e.g. 01010011 and 01010111, as preferred patterns, although in these two cases the longest run also occurs at the end.

Starting pattern. As indicated above, the effect of starting pattern interacts with rate, and the nature of this interaction on pattern description is shown in Figure 4. In Figure 4*a*, the use of the preferred starting pattern as the description is very great when the sequence is started at that pattern, and is much less when the sequence is started with the nonpreferred pattern. At a rate of 4·0 elements per sec., this effect begins to change, and at the highest rate there is no difference in use of the preferred starting pattern due to the actual starting pattern.

A similar sort of interaction is obtained for the second preferred pattern (Fig. 4*b*), in that at low rates the second preference is used much more frequently when the sequence is started with the nonpreferred pattern rather than with the preferred pattern. Again, the effect has essentially disappeared at the highest rate.

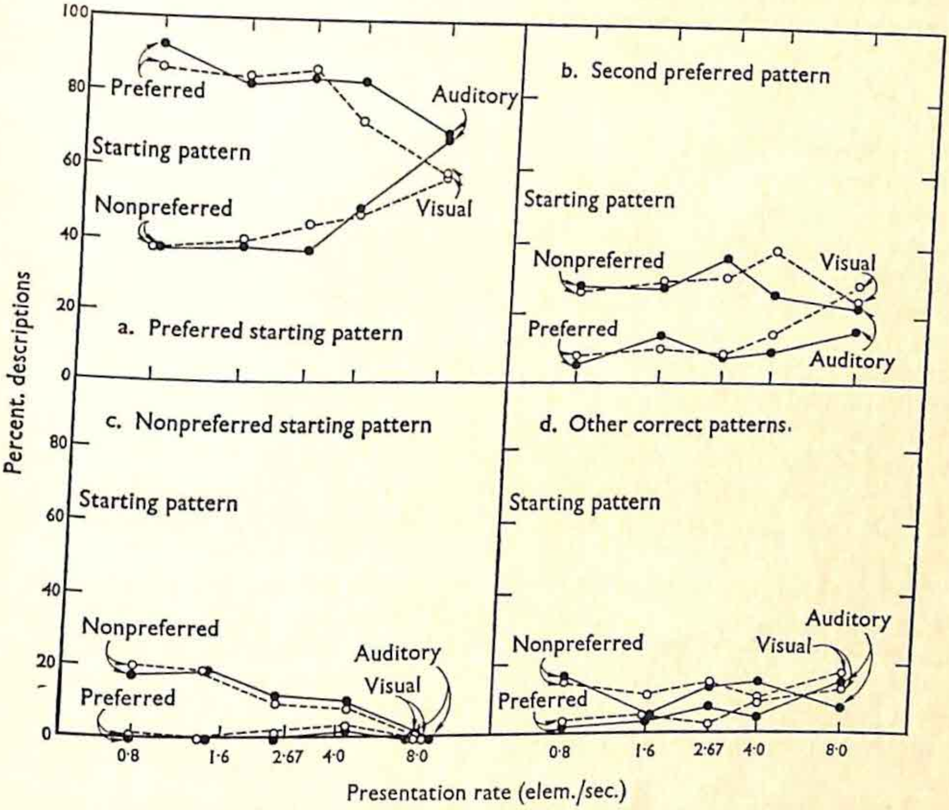
Still further, Figure 4*c* shows that the nonpreferred starting pattern is never used as a description unless the sequence is started with that pattern at low rates. At the highest rate, that pattern is never used as a description regardless of the starting pattern.

Figure 4*d* shows the use of other correct descriptions, with the same interaction once again. At low rates the effect of starting pattern is evident, with miscellaneous descriptions being used more frequently with nonpreferred starting patterns, while there is little or no difference at the higher rates.

Thus the starting pattern affects the description only at lower rates, at which the nonpreferred starting pattern increases the use of that pattern as a description along with an increased use of all other specific patterns except the preferred starting pattern.

Modality. Figure 4 also shows the effect of modality and its interaction with rate. The interaction in this case is again opposite that with starting pattern in that modality has essentially no effect on pattern description at low rates but does have some effect

FIGURE 4



Percentage of correct descriptions of different types as a function of rate of presentation of the pattern, for both modalities and both starting patterns. (Each plotted point is the arithmetic mean from the 10 different basic patterns. The particular pattern descriptions for each basic pattern are shown in Table II.)

at the higher rates. These effects at the higher rates are that auditory stimuli produce a higher average use of the preferred starting pattern, and a lower use of the second preferred and other patterns. There is no use of the nonpreferred starting pattern with either modality.

Pattern uncertainty. While use of uncertainty or variability of pattern description is not too valuable in the present study, one effect can be noted from the preceding effects of starting pattern on description as a function of rate. The general effect of starting pattern is to produce high variability of description with low rate and nonpreferred starting patterns, and very low variability with the preferred starting pattern, since the starting pattern is nearly always used as the description in this case. As rate increases, however, the descriptions become more alike, with uncertainty increasing for the preferred starting pattern and decreasing for the nonpreferred starting pattern.

DISCUSSION

The most consistent and striking result from this study is that both the starting pattern and the modality interact strongly with rate of pattern presentation in their effects on pattern difficulty and on pattern organization, but in entirely different ways. Whatever differential effects exist at low rates are due to differences in starting pattern; and whatever differential effects exists at high rates are due to differences in modality. It therefore seems parsimoniously reasonable to assume that two separate processes are involved in the experiencing of temporal pattern.

Pattern perception

At higher rates of stimulus presentation, certainly at rates above two elements per sec., it is most reasonable to talk about temporal pattern perception. At these rates the phenomenal experience is not one of a series of individual elements but of an *integrated* sequence, with the successive identical patterns having very definite beginnings and endings. When a sequence is first started at these higher rates, it appears as an unintegrated series which has no apparent organization, but then it fairly suddenly changes to one with a quite definite pattern.

The pattern organization at these higher rates is *phenomenally compelling*, in that once heard as a particular pattern it is almost impossible for it to be heard any other way. It is thus quite understandable that how the sequence is started has very little effect on the organization.

In a related fashion, the experience is *immediate*, not in the sense that no time is required for the organization to appear, but rather that the organization, once it has appeared, is a direct consequence of the physical stimulus. For this reason it is understandable that the modality should affect the organization and its difficulty, since the rapid timing relations will be intimately affected by the modality. The more precise timing allowed by the auditory modality is in keeping with the essential nature of the modality and with earlier results showing more precise timing relations with audition than with vision (Klemmer, 1967). We would expect even poorer performance if, for example, the stimuli were presented by touch.

Some reservation must be held concerning the result that auditory presentation leads to more accurate perception at higher speeds, even though the result is consistent with that of Klemmer. The reservation is due to the fact that almost certainly the inherent discriminability between the two binary events will have some effect on the accuracy of pattern perception (see e.g. Miller and Heise, 1950; Heise and Miller, 1951). Our intent in selecting the particular stimuli used was to have highly discriminable tones and lights, so that discriminability *per se* would not be a factor. But it is possible that the discriminability did not remain high and equal at the faster rates. Nevertheless, we do not think this is a factor, because the individual elements were easily perceived as separated from succeeding elements, for both modalities, and even at the fastest speed used.

Lastly, at these higher rates the role of the observer is relatively *passive*, and necessarily so. At the higher rates the elements occur too rapidly in succession for the subject to do anything except listen or look, and he certainly cannot make an overt motor response to each individual element. In work preliminary to the Royer and Garner study, in which a rate of two elements per sec. was used, these authors attempted to require responding to the sequence of sounds at the beginning of the sequence and found that the response is completely disruptive and prevents any perceptual organization from occurring. However, if the sequence is heard long enough for it to become an organized pattern, responding to the pattern at this rate is quite easy, and Royer's (1967) later study showed that even higher response rates are possible.

Pattern learning

At the lower rates of stimulus presentation it is more reasonable to talk about temporal pattern learning. At these lower rates the phenomenal experience is one of an *unintegrated* series of single elements, one which becomes a series of patterns only in a quite abstract sense.

The experience is in no sense compelling, nor is it direct and immmediate. Rather, the pattern is derived from the stimulus, in the sense that it can be and is recoded into another form. Thus the complete lack of effect of modality at these lower rates is quite reasonable, since it is the derived stimulus rather than the direct stimulus which is patterned.

The process at these lower rates is an *intellectualized* one, in which the observer tests hypotheses, forms new ones, and generally attempts to force organization into the sequence. The term rationalized perception is intuitively appealing for this process. With sufficiently complex patterns this intellectualized activity is not necessarily an advantage, since wrong hypotheses may be formed with a resultant greater organization time. Thus the effect that starting pattern has on this process is quite reasonable, because the subject attempts an immediate organization, as found in the Garner and Gottwald (1967) study, rather than waiting for the organization to appear, and it is thus readily interfered with.

Lastly, at the lower rates the role of the observer is much more *active*, both in the above sense of an intellectualized process and in the sense of his using more overt responding. The immediate use of responding does not interfere with the learning of a pattern, as Garner and Gottwald also found, nor does it facilitate learning. Thus the fact that the process is an active one at lower rates is not necessarily an advantage.

In summary, then, the perception of temporal pattern which occurs at the faster rates is one of an integrated sequence, is phenomenally compelling and immediate, and is a relatively passive process for the observer. On the other hand, the learning of temporal pattern is of a succession of single elements, and is a derived, recoded and intellectualized process in which the observer is much more active.

This research was supported by Grant MH 11062 from the National Institute of Mental Health to Johns Hopkins University, while the senior author was still on that university's staff. Preparation of the report was done while the senior author was a visiting scientist at M.R.C. Applied Psychology Research Unit, Cambridge.

REFERENCES

CLEMENT, D. E. (1964). Uncertainty and latency of verbal naming responses as correlates of pattern goodness. *J. verb. Learn. verb. Behav.*, **3**, 150–7.

GARNER, W. R. (1962). *Uncertainty and Structure as Psychological Concepts.* New York: Wiley.

GARNER, W. R., and CLEMENT, D. E. (1963). Goodness of pattern and pattern uncertainty. *J. verb. Learn. verb. Behav.*, **2**, 446–52.

GARNER, W. R., and GOTTWALD, R. L. (1967). Some perceptual factors in the learning of sequential patterns of binary events. *J. verb. Learn. verb. Behav*, **6**, 582–9.

HANDEL, S., and GARNER, W. R. (1966). The structure of visual pattern associates and pattern goodness. *Perception Psychophysics*, **1**, 33–8.

HEISE, G. A., and MILLER, G. A. (1951). An experimental study of auditory patterns. *Amer. J. Psychol.*, **64**, 68–77.

KLEMMER, E. T. (1957). Rhythmic disturbances in a simple visual-motor task. *Amer. J. Psychol.*, **70**, 56–63.

KLEMMER, E. T. (1967). Sequences of responses to signals encoded in time only. *Acta Psychologica*, **27**, 197–203.

MILLER, G. A., and HEISE, G. A. (1950). The trill threshold. *J. Acoust. Soc. Amer.*, **22**, 637–8.

ROYER, F. L. (1967). Sequential complexity and motor response rates. *J. exp. Psychol.*, **74**, 199–202.

ROYER, F. L., and GARNER, W. R. (1966). Response uncertainty and perceptual difficulty of auditory temporal patterns. *Perception Psychophysics*, **1**, 41–7.

Manuscript received 11th November, 1967.

FREQUENCY AND RECENCY EFFECTS ON RESPONSE BIASING IN A PSEUDO-PERCEPTUAL TASK

BY

NEIL McK. AGNEW, SANDRA PYKE and VERA PEZER
From York University, Toronto, Canada

The pseudo-recognition technique provides an ingenious method of evaluating the influence of response biasing variables. This study tests the effects of recency on low and medium frequency words using the pseudo-recognition procedure. It was found that under short recency conditions the thresholds of medium and low frequency words did not differ, while under longer recency conditions the low frequency words had significantly lower thresholds than medium frequency words.

INTRODUCTION

The question of whether variables such as frequency, recency or emotionality of stimulus materials affect seeing (perception) or saying (response) has received the attention of several investigators. Neisser (1954) found that preliminary presentation facilitated recognition of words but did not facilitate recognition of their homonyms. It should be noted that frequency was uncontrolled in this study. Neisser concluded that set or recency aided recognition but did not necessarily facilitate responses. However, a well known study involving pseudo-recognition techniques (Goldiamond and Hawkins, 1958) indicated that responses were biased by pre-exposure of the stimuli in the absence of perceptual cues in the testing situation. Using nonsense syllables, these investigators manipulated frequency, and their results suggest that the significantly lower thresholds for high frequency as compared with low frequency stimuli can be explained entirely in terms of response bias. A more recent study (Zajonc and Nieuwenhuyse, 1964) contains findings which indicate that the extent to which response bias affects results is a function of both frequency and criterion level for threshold. The more stringent the criterion and the lower the frequency, the less the role played by the response biasing factor.

Although the question of whether variables like frequency and value affect the motor response or the perceptual process was the primary concern of the investigations discussed above, the pseudo-threshold technique (Goldiamond and Hawkins, 1958; Newbigging, 1960) offers an interesting approach to the study of response biasing *per se* quite apart from any relation to perception. In the study presented here, the effects of frequency and recency on the strength of the response is determined by employing a pseudo-threshold technique. The hypothesis under test is that recent experience with a stimulus regardless of its frequency level in the population norms will serve to increase the strength of a congruent response. If the stimulus has a low population frequency, recent experience with the stimulus may increase the probability of this stimulus being emitted as a response up to or beyond that of a stimulus more frequently but not recently experienced. The specific prediction was that subjects would obtain lower pseudo-recognition thresholds for recently experienced stimuli as compared to stimuli not so recently experienced, and that this effect may be greatest with low frequency material, due to a relative ceiling effect on high frequency material.

METHOD

Subjects

The subjects for this study were 40 introductory psychology students (between the ages of 18 and 30 years) attending the University of Saskatchewan. The sample was divided into four equal groups each consisting of five males and five females.

Apparatus

A Pharmaceutical Research three-channel tachistoscope was employed. The blank channel was the pre-exposure field and contained a cross which served as a focal point for the subject. Its illumination level was 10·75 lumens. The channel called Stimulus I was the exposure field. It contained part of a Rorschach ink blot (Plate 1), and the illumination level was 12·75 lumens. Both the pre-exposure and exposure fields were 5 in. × 7 in.

A Hunter timer in circuit with the tachistoscope was set so that the pre-exposure field was constantly illuminated and the subject could see the cross. When the experimenter pressed a footswitch connected to the Hunter timer, the pre-exposure field went off and the exposure field was illuminated so that the subject was able to see the Rorschach blot for 0·02 sec.

Additional apparatus included a Tandberg tape recorder, Armaco earphones and a tape recording of the Maudsley Vigilance Test.

Materials

Stimulus materials consisted of the 20 words presented in Table I. The number in parenthesis represents the frequency of the word (Thorndike and Lorge, 1944). Low frequency words are listed in terms of the number of occurrences in four million words and the figures for medium frequency words represent the number of occurrences in one million words. The words were randomly placed in two lists.

TABLE I

STIMULUS MATERIALS

List A				*List B*			
Medium frequency per 1,000,000		*Low frequency per* 4,000,000		*Medium frequency per* 1,000,000		*Low frequency per* 4,000,000	
degrade	(8)	opaline	(11)	agitate	(9)	biplane	(10)
eclipse	(8)	seepage	(8)	tedious	(9)	curable	(8)
orderly	(9)	veining	(9)	shorten	(8)	tussock	(11)
magnify	(9)	verbose	(9)	essence	(8)	regress	(5)
screech	(8)	amnesia	(5)	calcium	(9)	atavism	(5)

Each word was typed (pica type) in capital letters on seven 5 in. × 7 in. white index cards and placed into two decks (list A and list B) each consisting of 70 cards. Responses were recorded on specially prepared data sheets consisting of 29 rows and 20 columns with one of the 20 stimulus words heading each column.

Procedure

All words were presented to all subjects but the order of list presentation was counterbalanced. Similarly, two levels of recency were used for all subjects and order of presentation of the recency levels was counterbalanced. A 30 sec. delay between pre-exposure of words and the pseudo-threshold determinations and a 6 min. delay were the two recency levels employed. Thus, four groups of subjects served under the various combinations of list and recency order but they all received the same stimuli and recency levels. Group I was tested with list A stimuli under short delay conditions followed by list B stimuli with long delay. Group II was tested with list A stimuli under long delay conditions followed by list B with short delay. Groups III and IV were identical to the groups described above except that the order of list presentation was reversed. The assignment of subjects to groups was alternated.

The procedure was as follows. The experimenter showed each subject the appropriate word on cards one at a time (list A or list B depending upon the group to which the subject had been assigned) and instructed him to read aloud each word as it was presented. The appropriate deck was thoroughly shuffled for each subject so that word order was randomized. Each of the 10 words of the appropriate list, A or B, appeared seven times in the shuffled deck so the subjects were exposed to each of the 10 words seven times. The presentation of the 70 cards took approximately 5 min. There was then a 30 sec., or 6 min., delay, during which the subjects performed the Maudsley Vigilance task. This consisted of a tape recording of random digits read at a fast rate delivered through earphones. Subjects were instructed to keep track of the number of threes appearing on the tape. The purpose of this task was to reduce the opportunity for spontaneous rehearsal of the words. Next, each subject received instructions, concerning the pseudo-recognition procedure (which took approximately 2 min.). The actual instructions for the first list were as follows: "I am going to present to you, one at a time, on the tachistoscope, the words you pronounced earlier from the deck of cards. If you look into the eye piece you will see an X. The word will appear where you see the X now. Each word will be presented for only a very brief instant, and you may not be able to tell exactly what the word is at first. However, after each presentation I want you to guess what the word was. Remember, even if you don't recognize the word, I want you to say what you think it is. Each word will be presented until you recognize it. I will tell you when you are correct, and then another word—or the same word, by random selection, will be shown. I will tell you when you have correctly identified the word, so if I don't say anything, you will know that you gave me an incorrect response and that you should try another word. I will say 'ready' before each exposure of the word. Any questions? Ready!"

On each recognition trial, part of a Rorschach ink blot (Plate I) was exposed for 0·02 sec. and subject's responses were recorded on the data sheet. When the subject said the word at the top of the column in which the experimenter was recording he was told he was correct and the experimenter began recording in the next column. If the subject failed to say the correct word on or before the 29th flash he was given a score of 30 for the word, and the experimenter moved to the next column. This entire procedure was repeated with the other deck of cards and delay conditions. At the completion of the experiment, a short questionnaire was administered to determine what the subjects concluded about the experimental procedure used.

The pseudo-recognition procedure for 10 words required a maximum of 20 min. Therefore, the average time delay between the pre-exposure recency condition and the pseudo-recognition trial for a given word was 15 min. for the 30 sec. condition and 20·5 min. for the 6 min. condition. These times are based on the average delay between the midpoint of the recency presentation and the midpoint of the pseudo-recognition procedure. The minimum delay between the pre-exposure of a word and the commencement of the pseudo-recognition trials was 2·5 min. (30 sec. plus 2 min. for instructions) under the 30 sec. recency condition, and 8 min. (6 min. plus 2 min. pseudo-recognition instructions) under the 6 min. recency condition. The minimum delay would occur in those cases where the last word under the pre-exposure recency condition happened to be the first word tested in the pseudo-recognition task.

RESULTS

The number of trials, including the trial on which a congruent response was obtained (the response was identical to the word at the top of the column), were totalled for each subject for each frequency and recency condition. These four scores were used in a three-way analysis of variance (Winer, 1962, p. 324) designed for cases when two variables are of the correlated or dependent type. The results are presented in Table II.

This analysis shows a significant over-all recency effect. Trials to criterion were significantly fewer under the 30 sec. condition as compared to the 6 min. condition. However, there was a significant recency by frequency interaction as outlined in Figure 1.* It is apparent that the recency effect is limited to the medium frequency words with the longer delay leading to higher thresholds. Conversely and surprisingly, low frequency words have slightly lower thresholds under long delay than short delay conditions.

The group by recency interaction is due to a practice effect. That is, all groups improve with the second recency condition regardless of duration (30 sec. or 6 min.).

Results of the post-experimental questionnaire indicated that most subjects (37 out of 40) believed that perception was involved in the experiment, so it appears that they did not "decode" the experimental procedure.

---

*This interaction has been replicated in a new series of experiments now being conducted.

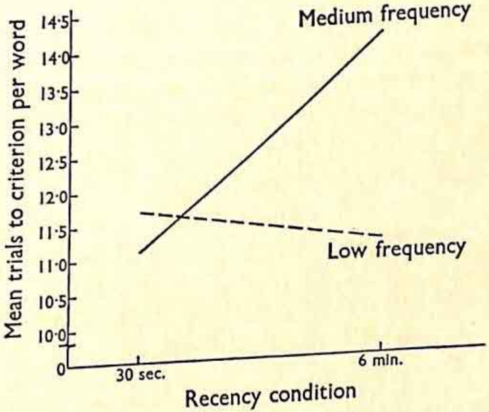
TABLE II

ANALYSIS OF VARIANCE

Source	*d.f.*	*M.S.*	*F*
Between subjects	39		
Groups (A)	3	2738·39	2·22
Error	36	1232·06	
Within subjects	120		
Recency (B)	1	3053·75	5·00*
A × B	3	2651·88	4·34*
Error	36	610·62	
Frequency (C)	1	2318·00	3·38
A × C	3	1572·39	2·29
Error	36	685·26	
B × C	1	5209·82	12·18†
A × B × C	3	304·42	<1
Error	36	427·57	
Total	159		

*$p < 0{\cdot}05$ † $p < 0{\cdot}01$

FIGURE 1



Mean trials to criterion as a function of frequency and recency.

DISCUSSION

The two main findings of this experiment are: (1) Under our short recency condition the pseudo-recognition thresholds of low frequency words do not differ significantly from those of medium frequency words, with the trend favouring the latter: (2) Under longer recency conditions the pseudo-recognition thresholds of low frequency words are (*a*) significantly lower than those of medium frequency words, and (*b*) slightly lower than those of low frequency words under short recency conditions.

Is there a single explanation that would account for both findings? If we consider the pseudo-recognition procedure as a form of free recall then we would expect frequency, rehearsal, and perhaps novelty as suspected sources of variance.

If frequency were the main variable we would expect significantly lower thresholds for medium frequency words under both delays, but in neither instance does this occur.

If rehearsal alone were the main variable we would expect two somewhat parallel distributions with slopes being a joint function of rehearsal and forgetting. We would hardly expect a marked interaction effect such as obtained, with low frequency words being significantly superior under the long delay condition. We are, therefore, looking for a variable with selective facilitative effects on low frequency words.

A facilitative factor favouring low frequency material could be the "novelty value" of the low frequency words. The effects of novelty on attention and perception are becoming increasingly well documented (Berlyne, 1960). The novely factor, favouring attention and concentration on low frequency words could account for the failure of medium frequency words to have a significantly lower threshold under the short recency conditions. Data from a study by Schultz presented by Noble (1963) indicates a negatively accelerated function between familiarity and recall with major increases in recall in evidence as you move from no familiarity to even small amounts of familiarity.

As we move from the short to long recency conditions we see the usual decreased availability of the response words for medium frequency material, but not for low frequency material. Not only is there no evidence of a forgetting function, but rather there is a slight increase in response availability moving from short to the long recency condition.

Although it is possible, it seems unlikely that the simple novelty effect would persist for such a time duration (average time between pre-exposure treatment and the pseudo-threshold recognition trials being approximately 20 min.). Some factor that counteracts the effects of forgetting over this time period seems operative. Though we attempted to prevent rehearsal with the Maudsley Vigilance task, as indicated by various authors (Brown, 1955; Hovland, 1951; Posner, 1967) it is extremely difficult to rule out rehearsal completely using the usual control procedures.

While only one of our subjects on the post-experimental open end questionnaire spontaneously noted that the Maudsley Vigilance task interfered with rehearsal, 14 of the 40 subjects commented in one way or another about the novelty value of some of the stimuli, and from personal experience the authors noted that the low frequency material did "intrude" during the vigilance task. Aside from rehearsal, we could invoke the concept of the "consolidation" of novel (arousing) stimuli, but at the present stage of knowledge this comes close to a tautology.

At present we favour an interpretation based on selective rehearsal of novel, or low frequency material to account for our results, with more opportunity for rehearsal occurring under the 6 min. delay than under the 30 sec. delay. These findings are of course relevant to Berlyne's proposal (1960) concerning the importance of "collative" variables in perception and recall, and underline the strong response biasing potential of recent, low frequency stimuli.

The results of an increasing number of studies appear to straddle some of the traditional, and perhaps arbitrary, boundaries between "perception," "short term memory," and "long term memory." Recently, Posner (1965) and Peterson (1966) have brought together some such studies and make intriguing attempts to integrate the findings.

This research was supported by Grant APT-60 from the National Research Council of Canada.

References

Berlyne, D. E. (1960). *Conflict, Arousal and Curiosity*. New York: McGraw-Hill.

Brown, J. (1955). *An Experimental Study of Immediate Memory*. Unpublished doctoral dissertation, Cambridge University.

GOLDIAMOND, I., and HAWKINS, W. F. (1958). Vexierversuch: the log relationship between word frequency and recognition obtained in the absence of stimulus words. *J. exp. Psychol.*, **56**, 457–63.

HOVLAND, C. I. (1951). Human learning and attention. In STEVENS, S. S. (Ed.). *Handbook of Experimental Psychology.* New York: Wiley, Pp. 613–89.

NEISSER, U. (1954). An experimental distinction between perceptual process and verbal response. *J. exp. Psychol.*, **47**, 399–402.

NEWBIGGING, P. L. (1960). Personal values and response strength of value-related words as a measure in a pseudo-perceptual task. *Canad. J. Psychol.*, **14**, 38–44.

NOBLE, C. E. (1963). Meaningfulness and familiarity. In COFER, C. N. (Ed.). *Verbal Behaviour and Learning.* New York: McGraw-Hill, Pp. 76–119.

PETERSON, L. R. (1966). Short term verbal memory and learning. *Psychol. Rev.*, **73**, 143–207.

POSNER, M. I. (1965). Memory and thought in human intellectual performance. *Brit. J. Psychol.*, **56**, 197–215.

POSNER, M. I. (1967). Short term memory systems in human information processing. *Acta Psychologica*, **27**, 267–84.

THORNDIKE, E. L., and LORGE, I. (1944). *The Teacher's Word Book of* 30,000 *words.* New York: Columbia Univ. Press.

WINER, B. J. (1962). *Statistical Principles in Experimental Design.* New York: McGraw-Hill.

ZAJONC, R. B., and NIEUWENHUYSE, B. (1964). Relationship between word frequency and recognition: perceptual process or response bias? *J. exp. Psychol.*, **67**, 276–85.

Manuscript received 17th July, 1967.

PUPIL SIZE AND PROBLEM SOLVING

BY

J. L. BRADSHAW*

From the Department of Psychology, University of Sheffield

Pupil size changes were monitored during the solution of various types of problems. A number of solution and response strategies were required of the subject. There was strong confirmation of the theory that this autonomic index can provide a sensitive measure of the fluctuating levels of attention and arousal, which are associated with the various aspects of information processing and response.

INTRODUCTION

Recent evidence supports the roll of arousal as an intervening variable in studies involving information processing tasks and changes in autonomic activity. Thus Hess (1964) found a relationship between changes in pupillary dilation and critical moments in solving mental arithmetic problems, as well as differences in overall problem difficulty. Kahneman and Beatty (1966), with short-term memory tasks, related pupillary changes to task difficulty and input and output stages. Glass (1966) unsuccessfully attempted to relate the degree of occipital α blocking to problem difficulty in multiplication tasks. Bradshaw (1967) used both cartoon and anagram material. He found that deliberately incomprehensible cartoon drawings tended to evoke greater pupillary dilation than ones which were understood, and that novel ones produced the same effect when compared with the re-presentation of familiar material. He found that with solved single-solution anagram problems, there was a peak in pupillary dilation at the moment of verbalized solution, followed by a rapid constriction. Unsolved and deliberately insoluble material caused a maintained relative dilation, as compared with the pupillary responses to solved problems, and to control abstract geometric stimuli.

The following two experiments extend these earlier studies by using single and multiple solution problems, immediate and stored response, which itself could be either verbal or a substitute motor activity, and problems varying in relative difficulty.

APPARATUS

All stimuli were prepared as slide transparencies, and checked for equal values of light transmission, using both a spot photometer and a light meter. However, the mode of construction of these stimuli, small black symbols transferred to a transparent background, had made such differences extremely unlikely. They were projected for 15 sec. each on to a back-projection screen, 12 in. × 16 in., 4 ft. from the subject's face. The latter was firmly positioned in an adjustable head holder. A Bolex 16 mm. reflex cine camera filmed the left eye of the subject, at a lens-to-eye distance of some 8 in. Both eyes were, however, unimpeded in their view of the stimulus screen. The camera was fitted with a 75 mm. telephoto lens, 5 mm. extension ring, and 3 dioptre supplementary lens. The complete eye, corner to corner, exactly filled one frame, reproducing therefore at about one-third life size. The camera was externally driven at 2·7 frames per sec., giving approximately 40 frames per stimulus exposure, and ⅓th sec. exposure per frame.

The light necessary to illuminate the eye came from the back-projection screen (100 ft.L.), and two small 15 watt (140 ft.L.) supplementary sources. At a distance of some 10 in. from the eye, each 60° from the direction of fixation, they provided tangential illumination from above and below the eye. The position of their reflected highlights afforded an accurate check on the direction of maintained fixation. All film records indicating deviations, eye movements or blinks were discarded from analysis.

---

* Present address: Department of Psychology, University of Otago, Dunedin, New Zealand.

The film used was Ilford HPS, rated at 650/800 ASA. However, by using a special prolonged dilute developing process it was exposed very satisfactorily at over 5,200/6,400 ASA, which is considerably faster than high speed infra-red film, and much more convenient.

The developed film was projected in strip form from a modified still projector, giving a very short wide throw. Projected directly on to graph paper divided into 1 in. and 1/10th in. squares, it was arranged so that a projected measurement of, say, 3·5 in. exactly equalled 3·5 mm. at the eye. Accurate focusing of the optical systems enabled a direct read-off of the pupillary diameters, the pupil contours being sharp at less than 1/10th in., or 0·1 mm. at the eye.

As the stimulus screen momentarily darkened immediately before and after each display, the absence of this reflected high-light on the film record aided calibration of temporal events. Twin-channel tape recording was also used, one channel providing an amplified record of the faint operation of the camera shutter, the other any ongoing events such as subject's responses, signals to the subject, or the sound of the stimulus change mechanism.

EXPERIMENT 1

This experiment was designed to compare the relative effects of easy and difficult rote arithmetic problems, and easy and difficult multiple-solution word game problems of the anagram type. Also to be compared were the effects of immediate verbalization upon attaining solutions, and the storing of responses until receipt of a signal. In this way it was hoped to separate out any independent or additive effects of verbalization and cognitive solution.

Material

There were in all 30 mental arithmetic and 30 word game problems. The former were division sums, with remainders, such that the answer was to be given to the nearest whole number. It was hoped in this way partly to control for differential knowledge, among the subjects, of multiplication tables. There were 15 easy problems, and 15 difficult, of the respective formats 146/5 and 256/17. The material word game consisted of a six-letter stimulus word, e.g. GROUND, CYCLIC, from which as many as possible words of three or more letters were to be constructed. A purely operational measure of difficulty was employed in that those problems where one or no solutions were found, were counted as difficult, and those with two or more as easy.

Procedure

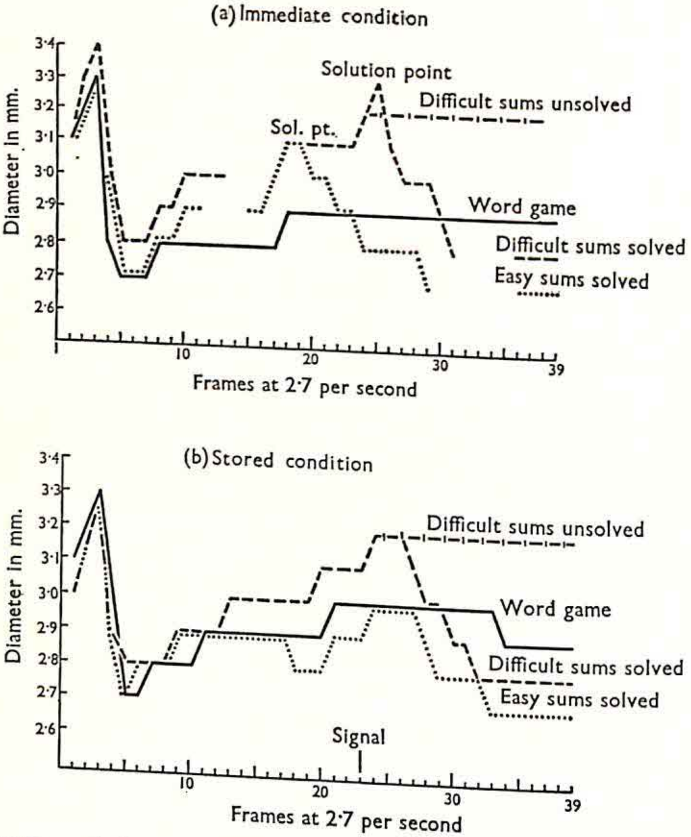
Six postgraduate students were used as the main subjects. They were given 15 min. adaptation time, followed by a considerable amount of prior practice. They were instructed to maintain fixation at all times upon a central fixation point on the screen. There were two conditions, an immediate-response condition, and a stored condition where the subject would respond, if able, with the solution upon hearing a buzzer signal. This occurred always at the same point in time, after about 9 sec. from stimulus onset. The material in each condition was otherwise similar. There were eight easy and seven difficult division sums in the first, immediate response, condition, with 15 word game problems, and seven easy and eight difficult sums in the stored condition, with the other 15 word game problems. All material was presented in randomized order.

Results

After measurement, the pupillary responses of each subject were calculated exactly as were the combined means of all subjects, given in Figure 1. With all stimuli in the stored condition, and with the word game material and unsolved arithmetic material in the immediate condition, straightforward averages were calculated for each photographic interval from stimulus onset; there were approximately 40 in a 15 sec. display. This was because responses occurred either at random intervals throughout, or not until the stored-response buzzer sounded. However, with the solved arithmetic problems in the immediate condition, there was a single solution which once attained and verbalized, closed the problem. Consequently averages were taken with shifted time axes, recentred upon the point of verbalized solution, for the central portions of the data.

In no case was there found any difference in pupillary response between those word game problems found easy, and those found difficult. Consequently, the data were combined into a single group mean curve, as given in Figure 1. However, with the arithmetic material, there was a very marked difference in dilation as between easy and difficult problems. Thus, in the immediate condition, four of the six subjects had significantly greater pupillary dilations to the difficult material, at $p < 0.001$, by the one-tailed sign test (Siegel, 1956). With the fifth subject the difference was also significant at $p < 0.05$, but the difference for the sixth subject was not significant. In the stored condition, all differences were again significant, four at $p < 0.001$, and two at $p < 0.05$. There was a rapid fall in diameter after solution, whereas unsolved material evoked a maintained and uniformly high plateau. With the immediate condition, where recentring about the point of actual and verbalized solution was possible, a solution peak over and above the level of the unsolved was frequently present.

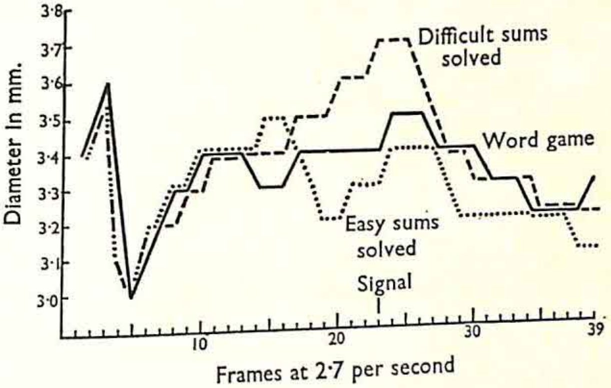
FIGURE 1



Mean scores of six subjects; (*a*) immediate and (*b*) stored conditions.

With the stored condition, where such recentring was not possible in the absence of a verbal response at the moment of solution, the high solution peak was absent. However, individually, with a number of subjects, a "real solution peak," followed by a drop before the further rise associated with the signal and the verbal response, was frequently and sometimes dramatically present. This was particularly evident with the easy material, which probably provided sufficient time for the two peaks to develop separately. This can be seen in Figure 2, which presents the individual responses of an additional subject.

FIGURE 2



Stored condition responses of a single subject, ceasing word-game solutions at signal.

The extra subject whose results are presented in Figure 2, was used in an attempt to determine whether a post-solution drop could be induced with the stored word-game material, exactly as with the solved arithmetic material, merely through instructions to cease attempting problem solution upon hearing the response signal. This indeed proved to be the case.

EXPERIMENT 2

The results of Experiment 1 confirmed that intrinsically more difficult single-solution problems lead to greater pupillary dilation. There was also some evidence to suggest that peaks in pupillary dilation could occur at the moment of problem solution, as well as to the act of verbalization. This second experiment was designed to compare further the effects of varying the cognitive load with different required processing strategies for solution, the effects of using verbal and substitute response activity, and of using single- and double-digit answer problems.

Material

Arithmetic material basically similar to that used in the first experiment was employed. There were four types of problem: those requiring single-digit answers, easy and difficult, and those requiring double-digit answers, easy and difficult. They were of the following formats respectively; 65/9, 147/17, 183/4, 414/18. The material was distributed among five experimental conditions, as designated below.

Procedure

There were five consecutive experimental conditions, as exemplified by the following instructions. The numbers in brackets refer to the number and type of the problems included.

1. Give your single-digit answer as soon as you can (eight easy, seven difficult single-digit problems).
2. Give your double-digit answer, digit-by-digit, as and when each one is attained (eight easy, seven difficult double-digit problems).
3. Give your double-digit answer as soon as you have ready both digits together (eight easy, seven difficult double-digit problems).
4. Do not verbally respond until the end of each display, giving then, if possible, the complete answer (eight easy double-digit problems).

5. Do not verbally respond until the end of each display, giving then, if possible, the complete answer. However, at the actual moment of solution, press your hand-held button-press (seven easy double-digit problems).

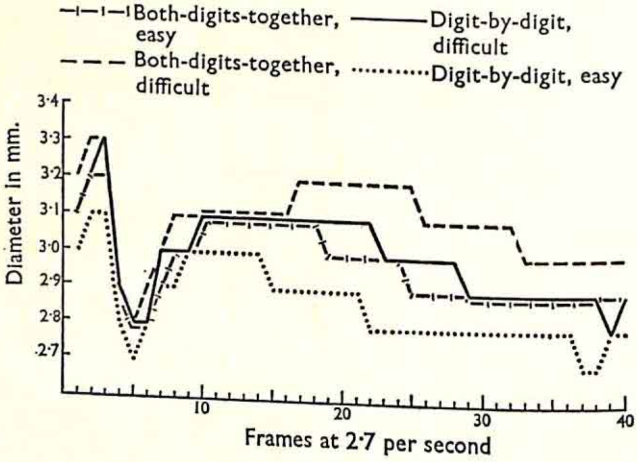
Seven post-graduate students were employed as subjects, and were given the usual prior adaptation and practice times. A counterbalanced order of presentation of easy and difficult problems was observed within each condition. A signal from a hand-held button-press recorded on the tape the subject's exact moment of solution.

Results

After processing, the responses of each of the seven subjects were calculated for all five conditions exactly as in the last experiment. In all the first three conditions where there were both easy and difficult problems, every subject evidenced significantly greater pupillary dilations with the more difficult material ($p < 0·001$ in 20 instances and $p < 0·008$ in one, using a one-tailed sign test).

When the double-digit, both-digits-together levels were compared with the operationally easier double-digit, digit-by-digit responses, the former were almost invariably associated with the greater levels of dilation, for both orders of problem difficulty ($p < 0·001$ in 10 instances, $p < 0·02$ in two, and N.S. in two, with a one-tailed sign test). Figure 3 provides the group mean scores, averaged in time, without recentring of the curves around solution points.

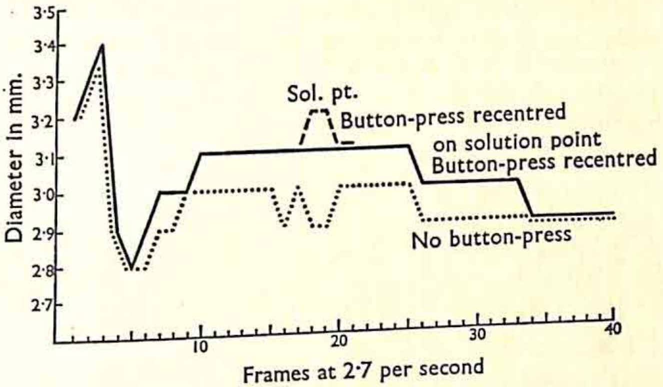
FIGURE 3



Mean scores of seven subjects; double-digit sums, "both-digits-together," and "digit-by-digit," easy and difficult.

The two last experimental conditions compared the effects of the presence, or absence, of a substitute motor (button-press) activity at the moment of problem solution, the subject verbalizing the answer between displays. Six of the seven subjects evidenced overall significantly greater levels of dilation to the button-press condition, as in Figure 4, each at $p < 0·001$ by the one-tailed sign test. The seventh difference was not significant. These differences, however, do not relate to the subsequent, immediately post-display levels, when the subject was verbalizing the stored response. Indeed, at this point only one of the seven subjects evidenced a significant difference ($p < 0·01$, *d.f.* 26, by a two-tailed t test); in all other cases the levels were exactly equal. Further, these selfsame levels at the moment of verbalization, in the button-press condition, were almost all significantly higher than those at the corresponding points in time of the double-digit, easy, both-digits-together problems, as in

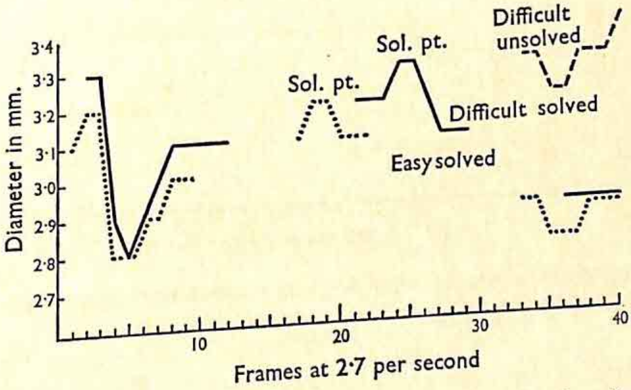
FIGURE 4



Mean scores of seven subjects: between-display verbalization, with and without button-press at solution.

Figure 4 ($p < 0{\cdot}01$ with three subjects, $p = 0{\cdot}02$ with two, $p < 0{\cdot}05$ with one, using a one-tailed t test, *d.f.* 22–28: the last was N.S.). The latter problems were exactly comparable to the button-press ones in difficulty. Inspection moreover of the individual sets of data showed that the recentred peak at the moment of button-pressing, in that condition, and shown superimposed in Figure 4, lay at the same height as another significant peak. This was the solution peak of the same, cognitively equivalent, easy, both-digits-together condition when recentred, as in Figure 5.

FIGURE 5



Mean scores of seven subjects, double-digit, "both-digits-together," easy and difficult, recentred.

The same was true also for both sets of declining, post-solution, measures. These points are also evident from the unrecentred versions of the data, in Figures 3 and 4.

Finally, the resetting or re-arousing effect of each new problem condition upon slowly and progressively declining pupillary measures, became apparent. This was demonstrated by a comparison of the mean pupillary scores associated with the first half of the problems, with those of the last half, in each of the last two conditions. All subjects showed overall significantly lower dilations for the latter half of each condition, in all 14 cases at $p < 0{\cdot}001$ by a one-tailed sign test.

Discussion

In general, it would seem that a peak in pupillary dilation occurs at the moment of solution with single-solution problems, and is followed by a sharp post-solution drop. That the latter is associated with "end-of-job" reduction in arousal is demonstrated by the presence of the drop, in Experiment 1, with multi-solution word game problems, when the subject was instructed to verbalize and thereafter terminate attention upon receipt of a signal. Greater pupillary diameters, moreover, are apparently associated with any form of increased cognitive load.

The role of arousal as a determining factor was further demonstrated in Experiment 2, in the comparison of early and late dilation levels in the last two conditions. These two conditions provided uniform material at the end of an experimental run, when the subject was likely to have settled down to the task. Thus all seven subjects, in both the last two conditions, exhibited significantly lower mean diameters during the second half of the experimental run. The levels were indeed reset temporally at other points when an occasional response known to be incorrect was made.

In Experiment 1, with the multi-solution word game material, the absence of solution peaks, the overall relatively low levels even with conditions of maximum response, and the lack of differences between such easy and difficult material, suggests two things. Either there is a fundamental difference between the two problem types, single and multi-solution, or that the act of verbalization does not account solely for the pupillary peak at solution. Certainly, multi-solution problems of this type are conceptually easier, and less defined in their methods or type of processing and solution than the rote transformations of single-solution arithmetic material.

That, however, the motor act of verbalization does contribute towards pupillary dilation is indicated by a number of facts. Thus, in Experiment 2, where the last two conditions were compared, one with and one without a button-press at the moment of solution, the former was associated with the higher pupillary levels. They were equal, in fact, to those others attained during the immediate verbalization of the solution. A substitute motor activity may perhaps therefore form a component in the associated response peak. However in this experiment it did not detract from the subsequent verbalization peak of the stored solution.

It is possible that some such factor as gratification or success at solving the task accounts for the dilation peak. If so, this could be tested in an ambiguous concept-formation task, with the experimenter randomly reinforcing decisions made by the subject with the comments "right," and "wrong."

This research was carried out at Sheffield University, as part of a Ph.D. investigation.

References

Bradshaw, J. (1967). Pupil size as a measure of arousal during information processing. *Nature*, **216** (5114), 515-6.

Glass, A. (1966). Comparison of the effect of hard and easy mental arithmetic upon blocking of the occipital α rhythm. *Quart. J. exp. Psychol.*, **17**, 142–51.

Hess, E. (1964). Pupil size in relation to mental activity during simple problem solving. *Science*, **143**, 1190–2.

Kahneman, D., and Beatty, J. (1966). Pupil diameter and load on memory. *Science*, **154**, 1583–5.

Siegel, S. (1956). *Nonparametric Statistics*. New York: McGraw-Hill.

Manuscript received 5th December, 1967.

ON CONRAD'S PREFIX AND GROUPING IN SHORT-TERM MEMORY

BY

HARRIS B. SAVIN

From Department of Psychology, University of Pennsylvania, Philadelphia

Requiring a subject to say a redundant prefix before he recalls a sequence of digits impairs his ability to recall that sequence. Evidence reported below suggests that the prefix functions exactly as an extra nonredundant digit—that the subject performs exactly as if the prefix had to be learned anew on each trial.

INTRODUCTION

Conrad (1958, 1960) complicated the usual digit-span experiment by requiring his subjects to say a completely redundant *O* just after hearing a sequence of digits and just before recalling it. Requiring the subjects to say the redundant prefix before recalling each sequence greatly impaired their ability to recall the sequences. Dallett (1964) replicated these results and, in addition, reported that the prefix had about the same effect upon overall accuracy as did an extra nonredundant digit. Conrad had previously (1958) considered the hypothesis that the prefix in fact functions just as an extra digit, as if it were not known to the subject in advance. He rejected this hypothesis, however, mainly because the first digit after the prefix (the first one the subject had to remember specifically for that trial) is recalled more accurately than is the second digit of a sequence when there is no prefix. Therefore, Conrad reasoned, the first digit after the prefix is not functionally like a second digit and hence the prefix is not functionally like a nonredundant first digit.

In the paper cited above, Dallett also reported results for the case in which the presented sequence always began with the same redundant prefix and the subjects were required to ignore it and to repeat only the subsequent digits. He found that this to-be-ignored prefix also impaired performance, by about as much as an extra nonredundant digit. He does not discuss serial position effects. In unpublished studies in this laboratory, involving much greater numbers of trials than did Dallett's experiment, slight but reliable differences have found between the various conditions that Dallett found to be equivalent with respect to overall performance. The differences fall into no coherent pattern; their direction depends upon rate of presentation of the digits. Also, serial-position effects depend upon both rate of presentation and the presence or absence of a prefix. Numerous small discrepancies aside, however, to a first approximation, these unpublished studies together with Dallett's suggest the hypothesis that the redundant prefix, whether it must be recalled or ignored, is functionally rather like a nonredundant extra digit that must be recalled. Rather than dismiss this hypothesis out of hand on account of the small discrepancies, it might be interesting to try to explain the discrepancies and save the hypothesis. This is the purpose of the present paper.

All the above-mentioned experiments, published and unpublished, suffer from a common weakness: grouping was not controlled. Subjects in digit-span experiments invariably say they group digits in various ways. Some grouping-patterns make for greater accuracy than others (Wickelgren, 1964), and, as will be shown below, the pattern of grouping has quite marked effects upon the shape of the serial-position curve. It would not be suprrising if, in the absence of any instructions to the contrary, subjects chose to group sequences differently depending upon whether there was or was not a redundant prefix. Conceivably, the prefix is in fact functionally just like an extra digit and the various small discrepancies mentioned above, which seem to rule out this possibility, were merely due to differences in preferred grouping-patterns in different experimental conditions. In Experiment I, therefore, the following four kinds of digit-span were compared: recall of nine digits, recall of eight digits, recall of a prefix and eight digits, and recall of eight digits which had been preceded by a redundant prefix that the subjects were required to ignore. In all cases, grouping was controlled by putting short pauses at the desired places in the presented sequences.

EXPERIMENT I

Procedure

Twenty-four subjects, students at the University of Pennsylvania, each heard 100 tape-recorded sequences, 25 of each of four kinds. On one fourth of the trials, each sequence consisted of nine digits, presented at the rate of 2 per sec. except for additional ½-sec. pauses after the third and sixth digits. The subjects tried to recall all nine digits. The condition will be called (9, 9). On another fourth of the trials, each sequence had nine digits, the first of which was always a 1; subjects were required to ignore the 1 and recall only the subsequent eight digits. This condition is called (P8, 8). Again, there were ½-sec. pauses after the third and sixth presented digit. On the remaining 50 trials, only eight digits were presented. On 25 of these, the subject simply recalled the eight digits (condition (8, 8)); on the other 25, he had to say a 1 just before recalling them (condition (8, P8)). On all 50 of these trials, there were ½-sec. pauses after the second and fifth digits. Otherwise, digits were presented at the rate of 2 per sec.

Each subject, then, served under four conditions. There are 4! = 24 possible orders of four conditions; each subject had them in a different order. Stimuli were tape-recorded and responses were spoken.

The digit 1 was used as the prefix because, in dialing any out-of-town telephone call from Philadelphia, one must prefix the telephone number with a 1; redundant 1-prefixes were therefore quite familiar to the subjects. 1 never occurred elsewhere in any sequence, nor did any other digit ever occur twice in a single sequence.

Subjects were given no instructions about how to group, but the most casual experimentation is enough to convince oneself that it is exceedingly hard to group sequences like these in any way other than the one defined by the pauses. The serial-position curves will provide further evidence that grouping was well-controlled.

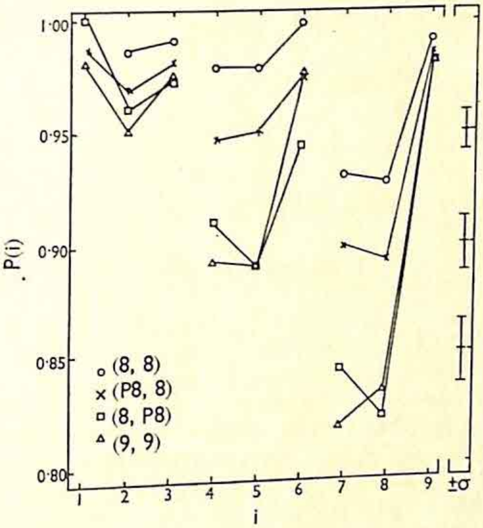
Results

The proportion of correctly recalled sequences is shown, for each of the four conditions, in Table I, and serial position curves are in Figure 1. Note that grouping of the digits is quite clearly reflected in the serial position curves (the last digit of a group is invariably relatively easy) and that, when grouping is controlled, the serial-position curves are quite similar for conditions (9, 9) and (8, P8).

There seems to be no satisfactory test of the statistical significance of the difference between these two curves. The problem is that they are not proper frequency distributions. Here and in numerous other studies in this laboratory, errors anywhere in the recall of a single sequence tend strongly to be accompanied by errors elsewhere in the same sequence. To facilitate a rough statistical assessment of the data, intervals of ±1 standard error of a proportion based upon 600 observations are shown on the right hand side of Figure 1, for several proportions in the range of interest. There are no obviously reliable differences between conditions (8, P8) and (9, 9). The data seem quite consistent with the view that, for one reason or another, the recalled prefix functions precisely as an extra nonredundant digit.

There is, however, one probably reliable difference between conditions (9, 9) and (8, P8): the prefix seems to be recalled slightly more accurately than is the first digit in (9, 9). Only two subjects ever omitted a prefix; one of them did so twice and the other

FIGURE 1



Proportion of correct digits at each serial position, Experiment I. Gaps in the curves indicate the places where there were pauses in the stimulus-sequences. For the purpose of these plots, the 8 serial positions in condition (8, 8) are considered positions 2 through 9, so that the pauses in the sitmulus-sequence are at the same abscissa values as in the other conditions. In condition (P8, 8), the first serial position is considered correct unless the subject said the prefix.

(Some readers have found this convention arbitrary. Fortunately, it does not matter. If all trials on which the subject is here considered in error on position 1 (condition (P8, 8)) were omitted altogether, the ordinates for positions 2–9 in the resulting graph would all be very slightly larger than they are here; none would increase by as much as 0·005.)

Vertical bars at the right-hand edge of the graph indicate a range of ± 1 standard error of the proportions whose ordinates are at the centre of the bars, for proportions estimated from 600 observations.

TABLE I

PROPORTION OF STRINGS CORRECTLY RECALLED IN EXPERIMENT I

	Condition	*Proportion correct**
A	(8, 8)	0·875 (0·012)
B	(P8, 8)	0·802 (0·03)
C	(8, P8)	0·682 (0·68)
D	(9, 9)	0·663

* Numbers in parentheses are two-tailed significance levels of differences between adjacent conditions, based upon sign tests of the within-subject differences.

only once. In condition (9, 9), the same two subjects and five more in addition failed to recall the first digit at least once. Conrad (1958) also remarks that prefixes are less likely to be omitted than nonredundant first digits.

Quite unexpectedly in the light of both Dallett's results and the author's unpublished ones cited above, the subjects were relatively successful at ignoring the prefix in condition (P8, 8). This condition was markedly easier than either (8, P8) or (9, 9), although still harder than (8, 8).

The serial-position curve for condition (P8, 8) (Fig. 1) suggests that, up until the first pause, the ignored prefix causes almost as much difficulty as an extra non-redundant digit, but, thereafter, it is relatively less troublesome—as if the subject were unable to ignore the prefix when he heard it but partially succeeded in eliminating its influence during the pause. Conceivably, had the pause been somewhat longer he would have been more completely successful in eliminating the prefix during it. Hence, Experiment II.

EXPERIMENT II

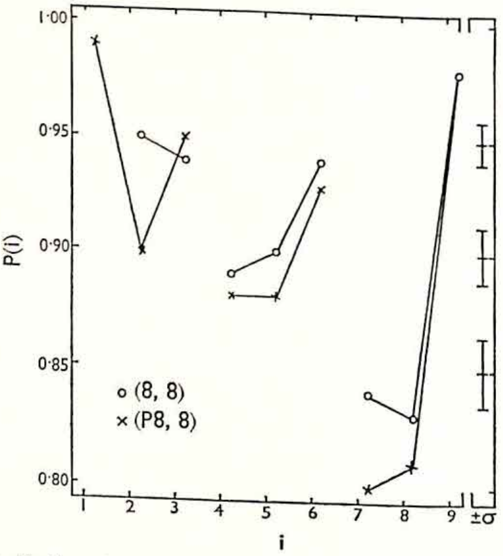
Procedure

The procedure was exactly that of Experiment I except as follows: Only conditions (P8, 8) and (8, 8) were used. The pauses in the stimulus-sequence were lengthened to 1·5 sec. each, and there were 12 subjects, each of whom heard 50 trials of each condition. Condition (P8, 8) was presented first to half the subjects and second to the rest.

Results

The proportions of correctly recalled sequences are 0·71 for condition (8, 8) and 0·64 for condition (P8, 8). The difference is significant ($p = 0{\cdot}038$) by a two-tailed sign test. Serial position curves are shown in Figure 2. Note that, for the six positions beyond the first pause, the distribution of errors over positions is much the same under the two conditions. Again, because the various serial positions are interdependent, no satisfactory statistical test is available, and the fact that the two curves do not cross cannot be readily interpreted.

FIGURE 2



Proportion of correct digits at each serial position, Experiment II. The same conventions are used here as in Figure 1.

Averaging over the last six serial positions, the proportion of correctly recalled digits was 0·895 in condition (8, 8) and 0·876 in condition (P8, 8). Seven subjects had higher proportions of correct digits in these positions in condition (8, 8) and four in condition (P8, 8). There was one tie.

The two conditions may differ slightly with respect to these six positions, but, in the light of Experiment I, it is reasonable to suppose that any small difference that may exist here would disappear entirely if the pauses in the stimulus were made still longer.

DISCUSSION

(1) *Recalled Prefixes*

A recalled prefix (e.g. the one in condition (8, P8)) appears in all respects indistinguishable from an extra nonredundant digit that must be remembered, except that the prefix is possibly a bit less likely to be forgotten.

The first question that arises in trying to understand this prefix effect is whether it arises during storage or during recall of the digits. The latter alternative seems untenable given the magnitude of the effect. If the prefix effect is some form of response interference, and if the prefix is functionally equivalent to an extra nonredundant digit, it follows that the ordinary digit-span must be limited only by response processes. It follows, in other words, that the subject has no particular difficulty in storing a long sequence of digits after a single hearing, and he can recall only about nine of them just because recall of the first digits interferes with retention of subsequent ones. Several recent studies (e.g. Anderson, 1960; Howe, 1965) have shown that there is a certain amount of this kind of response-interference in digit-span experiments, but nowhere near enough to sustain this conclusion. The prefix effect, therefore, cannot be wholly a matter of response-interference.

The alternative is that storing the prefix presents the subject with the same sort of problem as does storing a nonredundant digit. The consequences of this assumption depend crucially upon whether or not one assumes that short-term memory is a different system from long-term memory. If not—if one assumes only a single memory-system—then apparently insurmountable difficulties arise. Why should remembering (as opposed to recalling) the prefix impair the subject's performance any more than does remembering his name, his telephone number, and the immense store of other information that he carries with him?

If, on the other hand, one assumes a separate short-term memory, problems still remain but they are much less formidable. Assume, as must surely be the case, that the prefix is stored in long-term memory. (Subjects are never reminded to say it after they hear the instructions, and they almost never forget.) Further assume, however, that everything the subject recalls in experiments like these must also be stored in short-term memory and that short-term memory has a limited capacity. No obviously unreasonable consequences follow. On each trial, presumably, the subject must retrieve the prefix, from long-term memory and store it in short-term memory, where it requires just as much capacity as does a nonredundant digit. Even the fact—if it is a fact—that the prefix is omitted less often than a nonredundant digit can be explained. Because it is in long-term memory it cannot be misperceived, and there is ample time to store it in short-term memory. The only problem is that, as stated, the assumption that the prefix must be stored in short-term memory as well as in long-term memory seems impossibly *ad hoc*.

But this assumption, although not very well supported, is at least not totally unrelated to existing conceptions of short-term memory. There is good evidence that there is relatively little loss of information from short-term memory as long as rehearsal is permitted, and that any activity that precludes rehearsal results in very rapid forgetting (Peterson and Peterson, 1959). It is but a small additional step to the hypothesis that whatever is rehearsed must be stored in short-term memory. The precise nature of rehearsal is unclear, but it is often said to be quite similar to overt utterance except for the lack of muscular activity. If anything like this is true, then saying the prefix entails rehearsing it. If, in addition, rehearsing it requires that it be in short-term memory, then it must follow that a recalled prefix is functionally just like a nonredundant digit and that the subject is quite unable to take any advantage of the fact that he knows the prefix.

(2) *Ignored Prefixes*

Ignored prefixes are, if anything, even more perplexing than spoken ones. It is clear both from the subjects' performance and from introspection that, when the prefix is presented, as in these experiments, in the same rhythm as the following digits, subjects can neither ignore it altogether nor treat it simply as a "ready" signal. The best evidence is that they inadvertently say it almost as often as they fail to say the first digit in condition (9, 9).

When there is no pause in the presented sequence, it will be remembered, the ignored prefix affects overall performance (proportion of correctly recalled sequences) about as much as an extra nonredundant digit. Yet, it is somehow possible to use a pause which comes two digits after the prefix to largely mitigate its effects. Pursuing the speculations introduced above, it is tempting to argue that the prefix is initially stored just as a nonredundant digit (perhaps because the subject's attention-switching mechanisms are too slow for him to be able to ignore it and yet attend to another digit ½ sec. later) but, given a long enough pause in the middle of the sequence, it is possible to go back and "erase" the representation of the prefix. This explains the existing data but, for the present, it is not clear what further testable consequences it may have.

This research was supported by National Science Foundation grant NSF GS 857 to the University of Pennsylvania.

REFERENCES

ANDERSON, N. S. (1960). Post-stimulus cueing in immediate memory. *J. exp. Psychol.*, **60**, 216–21.

CONRAD, R. (1958). Accuracy of recall using keyset and telephone dial and the effect of a prefix digit. *J. appl. Psychol.*, **42**, 285–8.

CONRAD, R. (1960). Very brief delay of immediate recall. *Quart. J. exp. Psychol.*, **12**, 45–7.

DALLETT, K. M. (1964). Effects of a redundant prefix on immediate recall. *J. exp. Psychol.*, **67**, 296–8.

HOWE, M. J. A. (1965). Intralist differences in short-term memory. *Quart. J. exp. Psychol.*, **17**, 338–42.

PETERSON, L. R. and PETERSON, M. V. (1959). Short-term retention of individual verbal items. *J. exp. Psychol.*, **58**, 193–8.

WICKELGREN, W. A. (1964). Size of rehearsal group and short-term memory. *J. exp. Psychol.*, **68**, 413–9.

Manuscript received 24th October, 1967.

SEMANTIC DISTINCTIONS AND MEMORY FOR COMPLEX SENTENCES

BY

HERBERT H. CLARK and EVE V. CLARK

From the Department of Psychology, Carnegie-Mellon University and Department of General Linguistics, University of Edinburgh

This investigation studied what people remember in recalling complex sentences, whether it is certain semantic distinctions or merely transformational markers. After short intervals 24 subjects tried to recall sentences of six kinds which formed paraphrase sets: S_1 *before* S_2, S_1 *and then* S_2, *After* S_1 S_2, S_2 *after* S_1, S_2 *but first* S_1, and *Before* S_2 S_1. (S_1 and S_2 denote first and second clauses in temporal, not linguistic, order.) Subjects remembered the underlying sense of sentences with S_1–S_2 clause ordering better than those with S_2–S_1 clause ordering, regardless of transformational complexity. Subjects also showed a response bias, hence better verbatim recall, for sentences with subordinate clause second and for sentences with S_1–S_2 clause ordering. Sentence confusions indicated that subjects remembered three semantic distinctions: the temporal order, order of mention, and main-subordinate relation of the two described events. A theory of memory for marked and unmarked semantic distinctions was used to account for the results.

INTRODUCTION

What do people remember when they remember sentences? It is a common observation that people, when asked to remember a sentence verbatim, often err by recalling a paraphrase of that sentence. Mehler's (1963) subjects, for example, often confused active and passive sentences in recall, but did remember the underlying sense of the sentences. Miller's (1962), and Mehler's (1963), answer to the above question was that subjects remember the underlying kernel string and transformation markers of a sentence independently, and that they are very likely to forget the markers alone. Fillenbaum's (1966) subjects, on the other hand, made lexical, not transformational, mistakes while preserving the underlying sense. They replaced *not open* with *closed*, *dead* with *not alive*, and so on. As Fillenbaum argued, people seem to retain semantic information which is not bound to the linguistic form in which it was originally expressed (cf. Sachs, 1967).

The purpose of the present study was to show that memory for sentences has a semantic, rather than a syntactic, basis. Transformational changes, which previous investigators have studied, also entail semantic changes (Katz and Postal, 1964). Do people remember the transformational markers, then, or their correlated semantic distinctions? Actives and passives, for instance, are used in distinct ways, even though in the most general sense they are synonymous: speakers place what they are talking about at the time as the grammatical subject in both actives and passives (Clark, 1965; Halliday, 1967*a*, 1967*b*; Johnson-Laird, 1968). To illustrate, Morton (1966) has pointed out that

I saw the house. The house was built by the man.

and

I saw the man. The man built the house.

are both natural two-sentence sequences, although

I saw the house. The man built the house.

and

I saw the man. The house was built by the man.

both seem somewhat odd. The first sentence in each pair introduces what the reader expects to be the "theme" of the second sentence, but in the second two sequences the subject of the second sentence does not fulfil expectations. Thus subjects recalling a passive sentence might remember the semantic information that it was the recipient of the action that was talked about and not the actor, and this information would lead them to reconstruct a passive sentence. Because transformations generally entail semantic changes (e.g. questioning, negation, changes in emphasis, etc.), the experiments of Mehler (1963), Miller and McKean (1964), Gough (1965), Savin and Perchonock (1965), and others are difficult to interpret in strictly syntactic terms.

To demonstrate that people remember semantic distinctions that are not correlated with transformational changes, we chose sentences which contain a main clause and a subordinate clause (using *before* or *after*) and which describe two temporally ordered events. These sentences were selected because of two observations the second author had made in her study of the spontaneous speech of 3½-year-old Scots children (attending the nursery school of the SSRC Cognition Project, University of Edinburgh). First, in describing two related events with two clauses, the children linked the two clauses by intonation and almost always placed them in chronological order. This they did before—and after—they began to mark co-ordinate and subordinate clauses with the appropriate conjunctions. Second, the children at first almost always placed subordinate clauses after the main clauses. These two facts, taken as organizing principles, can be independently manipulated in sentences with *before*- and *after*-clauses. Sentences with *after*-clauses first and *before*-clauses second satisfy the first principle—that events are mentioned in chronological order; but sentences with *before*-clauses first and *after*-clauses second fail to do so.

The opposition of the two principles in *before*- and *after*-clauses allows the experimental verification of a semantic explanation of memory. If the first principle is assumed to be basic to memory, then sentences with *before*-clauses second and *after*-clauses first should be easier to remember than their alternatives. If the second principle of subordinate clause ordering is assumed to be basic to memory, then sentences with *before*- and *after*-clauses second should be easier to remember. Evidence for the second principle alone would support Miller's (1962) transformational complexity model, since a complex sentence with subordinate clause first is presumably the transformation of a complex sentence with the same subordinate clause second. Evidence for both principles, however, would support the semantic explanation of memory as the most general explanation.

METHOD

Sentences. Sets of six paraphrases, six synonymous complex sentences, were constructed from pairs of simple sentences. To illustrate, the two simple sentences, *The boy tooted the horn* and *The boy swiped the cabbages*, formed the basis of the following paraphrase set:

(1) S_1 *before* S_2: *He tooted the horn before he swiped the cabbages.*
(2) S_1 *and then* S_2: *He tooted the horn and then he swiped the cabbages.*
(3) *After* S_1 S_2: *After he tooted the horn he swiped the cabbages.*
(4) S_2 *after* S_1: *He swiped the cabbages after he tooted the horn.*
(5) S_2 *but first* S_1: *He swiped the cabbages but first he tooted the horn.*
(6) *Before* S_2 S_1: *Before he swiped the cabbages he tooted the horn.*

In the notation designating the linguistic form of the paraphrases, S_1 and S_2 refer to the first and second clauses in temporal, and not linguistic, order. In the first three sentences, S_1 precedes S_2 in the linguistic order, and in the last three S_1 follows S_2.

A total of 432 complex sentences, constituting 72 such paraphrase sets, were constructed from 72 pairs of simple sentences of the form, *The boy tooted the horn*, that is, sentences with transitive, past-tense verbs and simple subjects and objects preceded by *the*. The subjects of both simple sentences used to construct each paraphrase set were the same and were replaced in every complex sentence by the appropriate pronoun. The common subject was used as a noun cue for each complex sentence. Subjects were therefore presented with a noun cue and complex sentence, such as, *The boy. After he tooted the horn he swiped the cabbages.*

and were later given the noun cue alone and required to reproduce the sentence. The same noun cue was used for no more than two paraphrase sets. The simple sentences, all different from each other, had been randomly selected from sentences other subjects had composed for Clark (1965). The two simple sentences of each paraphrase set were assigned their temporal order randomly.

The use of noun cues as prompts has several advantages. First, it is natural. In everyday conversation one might ask someone, "What was it you just said about the boy?"; but one would be unlikely to ask, "What was it you said just after you had said the word *after*?" or some other question based on syntactic information. Second, the use of noun cues does not bias the results towards either the syntactic or the semantic explanation of memory. The cues, on any prior grounds, do not predispose subjects to use any of the six syntactic forms more often than any other, for the cues elicit both clauses of each sentence equally and give no indication of the subordinate conjunction or its placement. The important result, according to both syntactic and semantic explanations, is how often each of the six forms is misrecalled as one of the others; the two explanations differ only as to which confusions they predict. It is equally possible, therefore, for the confusions to support either explanation.

Procedure. For study, each subject was given a deck of IBM cards containing 72 pairs of noun cues and sentences, one from each paraphrase set. The 72 cards of paired cues and sentences were divided into 12 blocks of six with blank divider cards. For recall, each subject was also given a second deck of IBM cards which was identical to the first except that it contained only the noun cues. The subject was to write in the missing sentences beside the appropriate noun cues.

After an oral signal, each subject studied the first noun cue and sentence for 10 sec. at a second signal he turned over that card and studied the second pair for 10 sec.; he continued in this manner through the sixth pair. He then turned to the first card in the deck with the noun cues alone, and attempted to recall the appropriate sentence for the first noun cue. After success or failure on the first cue, he proceeded at his own pace (without intervention from the experimenter) to the second cue, third cue, and so on, through the sixth cue. After the sixth recall, he waited for the experimenter to signal him to study the first pair of noun cues and sentences in the second block. Then the procedure was continued as before.

Twenty-four subjects, all introductory psychology students at Carnegie-Mellon University fulfilling a course requirement, were run in groups of six. They were told that this was an experiment designed to study their memory for sentences. It was stressed that they should try to remember the sentences verbatim, and if they could not remember all of a sentence they should write down as much of it as they could remember. They were given an example of a card with the noun cue and its associated sentence printed on it, and a demonstration of the experimental procedure. The sessions lasted about 45 min.

Experimental design. The sentences were presented in a carefully counter-balanced design to free the data from all important position and sequence effects. The aims of this design were (*a*) to present a subject with six sentences, one in each of the six forms in an unpredictable order, in each block of sentences he received; and (*b*) to present each of the six sentences of one paraphrase set to a different subject but in the same position within the same block. Thus instances from the 72 paraphrase sets, and hence the noun cues, were ordered in the same way for all subjects. This order was constrained so that a noun cue occurred only once within any block of six successive cards. To accomplish the counter-balancing, six study decks of 72 paired noun cues and sentences were constructed with the following two properties: (*a*) Within the first (and each successive) block of six IBM cards of each deck, a 6 × 6 Latin square was formed from the six linguistic forms, the six positions (within the block), and the six decks. (*b*) Within the first six blocks and the last six blocks in each deck, another 6 × 6 Latin square was formed from the six linguistic forms, the six positions within each block, and the six blocks. The subjects were not told of these constraints, and there was no evidence that they perceived any of them. Each of the six study decks was given to four subjects.

RESULTS

The main results are shown in the matrix of confusions in Table I. The head of each column gives the form with which subjects were presented, and the head of each row gives the linguistic form of their responses. Recalled sentences were tallied when their linguistic form was recognizable (*then* and *and* alone were counted as the conjunction *and then*) and when at least one of the two clauses was partly correct. This method of scoring was used since the temporal order and order of mention of two events were of greatest interest and since most verbatim errors within clauses resulted from substitution of synonyms for single words. Approximatly one third of the recalled sentences in the first twelve rows of Table I contained at least one lexical error, but the matrix of confusions for the errorless sentences closely matches the matrix shown.

TABLE I

Confusions among Six Linguistic Forms

Sentence recalled			*Sentence presented*						
			(1)	(2)	(3)	(4)	(5)	(6)	Subtotals
Temporally correct	(1)	S_1 before S_2	170	52	36	11	8	41	318
	(2)	S_1 and then S_2	34	134	95	11	7	9	290
	(3)	After S_1 S_2	23	40	83	10	1	9	166
	(4)	S_2 after S_1	6	5	15	155	66	60	307
	(5)	S_2 but first S_1	0	1	0	5	83	17	106
	(6)	Before S_2 S_1	1	0	1	8	14	52	76
		Subtotals	234	232	230	200	179	188	1,263
Temporally incorrect	(1′)	S_2 before S_1	3	4	3	23	30	29	92
	(2′)	S_2 and then S_1	1	3	5	14	20	12	55
	(3′)	After S_2 S_1	0	1	1	11	8	13	34
	(4′)	S_1 after S_2	11	8	6	6	3	3	37
	(5′)	S_1 but first S_2	3	2	3	3	4	2	17
	(6′)	Before S_1 S_2	3	4	2	0	3	1	13
		Subtotals	21	22	20	57	68	60	248
		Blanks	15	12	22	10	17	20	96
		Miscellaneous	18	22	16	21	24	20	121
		Totals	288	288	288	288	288	288	1,728

The responses in the first six rows of Table I are synonymous with the presented sentences. Those in the second six rows have errors in temporal order in that they describe the two events as happening in reverse of the presented sentences. In the first three rows the responses have an S_1–S_2 clause order; in the next three rows they have an S_2–S_1 clause order.

A semantic explanation of memory was clearly supported by the data. Subjects remembered the underlying sense of S_1–S_2 sentences more often than of S_2–S_1 sentences, regardless of the linguistic form of the sentences. This effect is seen in the column subtotals for the first six rows in Table I; the subtotals 234, 232, and 230 for S_1–S_2 sentences are consistently larger than 200, 179, and 188 for the S_2–S_1 sentences. By an analysis of variance, these six subtotals differed significantly, both with the subjects as the sampling variable ($F = 9{\cdot}63$, $d.f. = 5,\ 115$, $p < 0{\cdot}001$) and with paraphrase sets as the sampling variable ($F = 12{\cdot}32$, $d.f. = 5,\ 355$, $p < 0{\cdot}001$). By Duncan's multiple range test, each of the S_1–S_2 sentences was significantly better remembered at $p < 0{\cdot}01$ than each of the S_2–S_1 sentences.

Not only did subjects remember the underlying sense of S_1–S_2 sentences more often, but they made fewer errors of temporal order for S_1–S_2 sentences. This complementary effect is seen in the column subtotals for the second six rows; the subtotals 21, 22, and 20 are consistently smaller than 57, 68, and 60. On the other hand, the sums of the two subtotals in each column are approximately equal; i.e. subjects recalled at least part of the presented sentence equally well for the six linguistic forms. Therefore, subjects remembered the sense of S_2–S_1 sentences less often, not because they remembered fewer S_2–S_1 sentence fragments altogether, but because they made errors in the temporal order of the two events.

Two distinct response biases were found in recall—one towards S_1–S_2 sentences and the other towards sentences with subordinate clause second. The linguistic forms subjects preferred as responses are indicated in the row subtotals (the last column) in Table I. By a Friedman two-way analysis of variance by subjects, the first six row subtotals differ significantly ($\chi^3 = 58{\cdot}06$, $d.f. = 5$, $p < 0{\cdot}001$). First, subjects preferred S_1–S_2 sentences over S_2–S_1 sentences: they produced S_1 *before* S_2 more often than S_2 *after* S_1; *After* S_1 S_2 more often than *Before* S_2 S_1; and S_1 *and then* S_2 more often than S_2 *but first* S_1 (the second two differences are significant at $p < 0{\cdot}005$ by a sign test). Second, subjects preferred sentences with subordinate clause second over sentences with subordinate clause first: the differences between the two sentences with *before*-clauses and between the two with *after*-clauses were both significant (at $p < 0{\cdot}005$ by a sign test).

The sentences subjects recalled in literally correct form directly reflected this response bias. They recalled more S_1–S_2 sentences than S_2–S_1 sentences, and more sentences with subordinate clause second than sentences with subordinate clause first. The tally of literally correct recalls is found in the diagonal of the first six rows in Table I. By analysis of variance by subjects, these diagonal entries differed significantly ($F = 16{\cdot}69$, $d.f. = 5, 115$, $p < 0{\cdot}001$), and by Duncan's multiple range test, they fell into two groups: S_1 *before* S_2, S_2 *after* S_1, and S_1 *and then* S_2 were each significantly better recalled at ($p < 0{\cdot}01$) than each of the others.

So far we have discussed two effects, those of S_1–S_2 clause ordering and main-subordinate clause ordering. A third effect, immediately clear in the confusions in Table I, is the effect of order of mention. After two events were mentioned in a particular order in a presented sentence, they tended to be recalled in that same order in the recalled sentence. For example, the first three columns represent S_1–S_2 sentences, i.e. sentences in which S_1 was mentioned before S_2. The predominant confusions for such sentences were with other S_1–S_2 sentences, even when a temporal error had been made. Likewise, S_2–S_1 sentences were predominantly confused with other S_2–S_1 sentences.

Discussion

To review the main findings: subjects were better able to remember the sense of S_1–S_2 sentences than of S_2–S_1 sentences. Their accuracy for sense was not related to transformational complexity. Their preferences for mode of response, however, were. In expressing what they remembered, subjects showed a bias towards sentences with subordinate clause second, but also towards S_1–S_2 sentences. Their literally correct recalls—as opposed to their sense-preserving responses—were in direct agreement with this response bias. Furthermore, subjects often remembered only the order of mention of two events, confusing the S_1–S_2 sentences with each other and the S_2–S_1 sentences with each other.

The theory that subjects remember kernels plus transformation markers (Miller, 1962) cannot account for the present results. It does not predict the subjects' superior memory for S_1–S_2 sentences over S_2–S_1 sentences or their memory for order of mention of two events, although it does predict their superior memory for sentences with subordinate clause second (if these sentences are assumed to be transformationally simpler than sentences with subordinate clause first). It will be shown that this latter finding can be included within a more general theory of memory for semantic distinctions, a theory which also predicts the first two findings.

The theory we propose to account for these data rests on the concept of marked and unmarked semantic distinctions. Many semantic features, such as the singular-plural feature in English, consist of two forms—the syntactically simpler, unmarked form, and the more complex, marked form. Singular nouns are unmarked, while plural nouns are marked, usually by an added *s* (see Greenberg, 1966). Analogously, semantic distinctions in sentences can be marked or unmarked. The semantic-distinction theory states that: (*a*) in reading sentences, subjects extract as objects of memory certain semantic distinctions; (*b*) subjects suffer memory loss by forgetting that a distinction was marked or, less often, unmarked, or by forgetting particular distinctions altogether; and (*c*) when attempting to recall earlier material, subjects reconstruct sentences based on the semantic distinctions they remember.

Thus subjects can forget two things—the marking of a distinction, or the distinction itself. They will forget that a distinction was marked, however, more readily than that it was unmarked. To illustrate, subjects might be presented with *open* and *not closed* to recall under appropriate circumstances. We assume that forms with *not* are marked, and those without are unmarked. At times subjects would forget the open-closed distinction, and recall *not open* or *closed* instead of the presented words. But in making errors of marking, they would recall *not open* as *closed* more often than *closed* as *not open*. This asymmetry is indicated in Table II, a hypothetical matrix of confusions for marked and unmarked forms. If the two forms were confused with each other equally, the matrix would have equal diagonal and equal off-diagonal entries. A marked form, however, is mistakenly recalled as an unmarked form more often than the reverse. Compared to a symmetric matrix, the upper right cell is increased and the lower right cell decreased by these asymmetric marking errors. This effect will also produce an apparent response bias towards the unmarked form, as seen in the row totals (see also Marshall, in press).

TABLE II

HYPOTHETICAL CONFUSION MATRIX OF MARKED AND UNMARKED FORMS

Recalled form ..		*Presented form*	
		Unmarked	*Marked*
	Unmarked ..	highest	third highest
	Marked ..	lowest	second highest

In the present experiment there seemed to be three main semantic distinctions subjects extracted for memory: the temporal order, order of mention, and main-subordinate relation of the two described events. The three distinctions are made of the events themselves and can be independently realized in the six linguistic forms. In *Before he swiped the cabbages he tooted the horn*, the tooting happened first and the theft second, in temporal order. But the theft was mentioned first and the tooting second, in order of mention. Furthermore, the tooting in the main clause was fixed in time and the theft in the subordinate clause was placed in time relative to the tooting. The last distinction is the main-subordinate relation of the events. The three distinctions were important to the present subjects, since they preserved each of them, respectively, in 84, 88, and 74 per cent. of their responses (excluding blanks and miscellaneous responses).

Of the three distinctions, two occur in marked and unmarked form; each marking serves to relate one distinction to another. Order of mention can be marked. Usually, order of mention indicates temporal order directly. It becomes marked when something in the sentence indicates that the event mentioned first actually occurred *after* the event mentioned second. In the above example, the tooting happened before the theft, but was mentioned after the theft. Thus the sentence is marked for order of mention, as are all S_2–S_1 sentences. The main-subordinate relation can also be marked. The usual unmarked form is with the main event mentioned first. The distinction is marked in sentences which mention the subordinated event first, as in the above example and all other sentences with subordinate clause first. Marking for order of mention indicates a specific interaction between order of mention and temporal order, whereas marking for main-subordinate relation indicates an interaction between main-subordinate relation and order of mention. The six presented sentences are given in Table III with their appropriate markings.

TABLE III

MARKED AND UNMARKED SEMANTIC DISTINCTIONS

	Semantic distinctions	
Linguistic forms	*Order of mention*	*Main-subordinate relation*
(1) S_1 before S_2	unmarked	unmarked
(2) S_1 and then S_2	unmarked	(unmarked)
(3) After S_1 S_2	unmarked	marked
(4) S_2 after S_1	marked	unmarked
(5) S_2 but first S_1	marked	(unmarked)
(6) Before S_2 S_1	marked	marked

The theory proposed here fits the data well if we assume these three distinctions, two of which can be marked. Order of mention will be considered first. Table IV gives a classification of order of mention errors in a confusion matrix. The upper half-matrix resembles the hypothesized configuration in Table II, as it should. The lower half-matrix, containing temporal errors too, must be considered separately. If subjects had made symmetrical marking errors, this half-matrix also should be symmetrical about its main diagonal. But some subjects must have remembered order of mention while forgetting temporal order. When presented S_2–S_1 sentences, such

TABLE IV

MATRIX OF CONFUSIONS BETWEEN S_1–S_2 AND S_2–S_1 SENTENCES

Recalled form		*Presented form*		
		S_1–S_2 Sentences (unmarked)	*S_2–S_1 Sentences (marked)*	*Totals*
	S_1–S_2 Sentences (unmarked)	667	107	874
Temporally correct	S_2–S_1 Sentences (marked)	29	460	489
	Subtotals	696	567	1,263
	S_2–S_1 Sentences (unmarked)	21	160	181
Temporally incorrect	S_1–S_2 Sentences (marked)	42	25	67
	Subtotals	63	185	248

subjects had their choice of recalling a marked temporally correct sentence (the cell with 460) or an unmarked temporally incorrect sentence (the cell with 160). Most subjects, preferring the unmarked form, chose the latter temporally incorrect sentences. When presented S_1–S_2 sentences, however, such subjects had to choose between an unmarked temporally correct sentence (the cell with 667) and a marked temporally incorrect sentence (the cell with 42). Again most of them chose the unmarked form, increasing the number of temporally correct responses. The cell with 160 had been inflated, and the cell with 42 deflated, because of memory errors towards the unmarked form. This reasoning would explain the important finding that subjects remembered the underlying sense of S_1–S_2 sentences more often than of S_2–S_1 sentences.

TABLE V

MATRIX OF CONFUSIONS BETWEEN SENTENCES WITH SUBORDINATE CLAUSE SECOND AND SENTENCES WITH SUBORDINATE CLAUSE FIRST

Recalled form		*Presented form*		
		Main-subordinate order (unmarked)	*Subordinate-main order (marked)*	*Totals*
Without order of mention errors	Main-subordinate order (unmarked)	359	131	490
	Subordinate-main order (marked)	45	150	195
	Subtotals	404	281	685
With order of mention errors	Main-subordinate order (unmarked)	26	66	92
	Subordinate-main order (marked)	11	12	23
	Subtotals	37	78	115

The main-subordinate relation, with its errors and errors of marking, produces a matrix of confusions (Table V) very similar to that of order of mention. (The two co-ordinate sentences are disregarded for the moment.) The upper half of the matrix is as predicted. The lower half is explained with the same reasoning as before. That is, when subjects had forgotten the order of mention but not the main-subordinate relation of two events, they tended to produce the unmarked sentence in recall. This inflated the upper right cell (with the 66) in the lower matrix, as well as the upper left cell (with the 359) in the upper matrix.

It was the unmarked forms that subjects most often preferred as their mode of recall. S_1 *before* S_2, which was completely unmarked, was used most often. *Before* S_2 S_1, which was marked for both order of mention and main-subordinate relation, was used least often. The other two subordinated forms, both marked on one distinction, fell in between the two extremes. The unmarked co-ordinate form was used more often than the marked form. The present theory would predict these biases, just as it predicted the bias in the above hypothetical example (Table II). A tendency for a distinction to lose its marking makes the unmarked form more frequent.

The three semantic distinctions mentioned above, to be of use, must have independent linguistic justification. All three, of course, have obvious syntactic origins. The presented sentences have purely syntactic devices to distinguish first and second events in temporal order, first and second events in order of mention, and main and subordinated events. Which forms are unmarked and marked, however, is a more difficult question. When a distinction is unmarked in a sentence, that syntactic form is simpler or more primary. One traditional way linguists have decided on which form is primary is to rely on an impressionistic judgement of frequency of usage. As mentioned before, young Scots children talk about events in their order or occurrence and almost never reverse the order. It is easy to see why In telling stories and in talking about events around them, they indicate the order of events. The reason might be that they know it is important to do so, or that they necessarily recall events in the order in which they happened, or otherwise. Whatever the reason, before they have at their command the use of subordination and certain relation words, the only way they can indicate temporal order is through order of mention. Order of mention, therefore, would seem to indicate the temporal order of events to young children, and any deviation from this pattern must be syntactically marked. A similar argument holds for the main-subordinate relation. When these same Scots children first use subordinate clauses, they almost always place the subordinate clause after the main clause (see also Davis, 1937; and Heider and Heider, 1940). This ordering is a reasonable choice, for in it the fixed event becomes the event first mentioned—the theme of the sentence or what the speaker is talking about—while the subordinated event is mentioned second to indicate that it "modifies" the theme.

The marking of the main-subordinate relation has further justification in its formal linguistic analysis. Lees (1961) suggests that the subordinate clause is imbedded in kernel strings or sentences in place of time adverbials (such as *then*). The clause *after he ate*, according to this view, replaces the word *then* in (the underlying structure of) the sentence *He left then* to produce the complex embedded sentence *He left after he ate.* Halliday (1967*b*), however, points out that when a speaker places *then* at the beginning of a sentence, as in *Then he left*, he is marking *then* as the theme of the sentence. In *He left then*, on the other hand, *he* is the unmarked theme, for the subject of the verb is the unmarked theme in English. Similarly, we argue that *After he ate he left* is marked and *He left after he ate* unmarked. When an English speaker wants to make the subordinated event itself the theme of his sentence, he marks the sentence syntactically by inverting the main-subordinate clause order.

This research was supported in part by Public Health Service research grant MH-07722 from the National Institute of Mental Health. We thank David Bree and John R. Hayes for their helpful comments on the paper.

REFERENCES

CLARK, H. H. (1965). Some structural properties of simple active and passive sentences. *J. verb. Learn. verb. Behav.*, **4**, 365–70.

DAVIS, E. A. (1937). The development of linguistic skill in twins, singletons and singletons with siblings, and only children from the age of 5 to 10. *Univ. of Minnesota Child Welfare Monograph* XIV.

FILLENBAUM, S. (1966). Memory for gist: some relevant variables. *Language and Speech.* **9**, 217–27.

GOUGH, P. B. (1965). Grammatical transformations and speed of understanding. *J. verb. Learn verb. Behav.*, **4**, 107–11.

GREENBERG, J. H. (1966). *Language Universals*. Janua Linguarum. The Hague: Mouton.

HALLIDAY, M. A. K. (1967*a*). *Some Aspects of the Thematic Organization of the English Clause.* Santa Monica: The RAND Corporation (Memorandum RM-5224-PR).

HALLIDAY, M. A. K. (1967*b*). Notes on transitivity and theme in English, part II. *J. Linguistics*, **3**, 199–244.

HEIDER, F., and HEIDER, G. (1940). A comparison of sentence structure of deaf and hearing children. *Psychol. Monograph* **52**, 42–103.

JOHNSON-LAIRD, P. N. (1968). The interpretation of the passive voice. *Quart. J. exp. Psychol.*, **20**, 69–73.

KATZ, J. J., and POSTAL, P. M. (1964). *An Integrated Theory of Linguistic Descriptions.* Cambridge, Mass.: M.I.T. Press.

LEES, R. B. (1961). Grammatical analysis of the English comparative construction. *Word*, **17**, 171–85.

MARSHALL, J. C. (in press). Psychological aspects of semantic structure. In MEETHAM, R. (Ed.) *Encyclopedia of Linguistics, Information and Control.* London: Pergamon.

MEHLER, J. (1963). Some effects of grammatical transformations on the recall of English sentences. *J. verb. Learn. verb. Behav.*, **2**, 346–51.

MILLER, G. A. (1962). Some psychological studies of grammar. *Amer. Psychol.*, **17**, 748–62.

MILLER, G. A., and MCKEAN, K. (1964). A chronometric study of some relations between sentences. *Quart. J. exp. Psychol.*, **16**, 297–308.

MORTON, J. (1966). Discussion of 'On hearing sentences'. In LYONS, J. and WALES, R. J. (Eds.) *Psycholinguistics Papers.* Edinburgh University Press, 10–23.

SACHS, J. (1967). Recognition memory for syntactic and semantic aspects of connected discourse. *Percept. and Psychophys.*, **2**, 437–42.

SAVIN, H. B., and PERCHONOCK, E. (1965). Grammatical structure and the immediate recall of English sentences. *J. verb. Learn. verb. Behav.*, **4**, 348–53.

Manuscript received 18th November, 1967.

SELECTIVE ATTENTION AND CEREBRAL DOMINANCE IN PERCEIVING AND RESPONDING TO SPEECH MESSAGES

BY

ANNE TREISMAN and GINA GEFFEN

From Medical Research Council Psycholinguistic Research Unit, Institute of Experimental Psychology, University of Oxford

The effect of attention on cerebral dominance and the asymmetry between left and right ears was investigated using a selective listening task. Right handed subjects were presented with simultaneous dichotic speech messages; they shadowed one message in either the right or left ear and at the same time tapped with either the right or the left hand when they heard a specified target word in either message. The ear asymmetry was shown only when subjects' attention was focused on some other aspect of the task: they tapped to more targets in the right ear, but only when these came in the non-shadowed message; they made more shadowing errors with the left ear message, but chiefly for non-target words. The verbal response of shadowing showed the right ear dominance more clearly than the manual response of tapping. Tapping with the left hand interfered more with shadowing than tapping with the right hand, but there was little correlation between the degree of hand and of ear asymmetry over individual subjects. The results support the idea that the right ear dominance is primarily a quantitative difference in the distribution of attention to left and right ear inputs reaching the left hemisphere speech areas. This affects both the efficiency of speech perception and the degree of response competition between simultaneous verbal and manual responses.

INTRODUCTION

The relations between cerebral dominance, speech functions and handedness have aroused renewed interest in the last few years (Zangwill, 1960; Kimura, 1961; Milner, 1962; Benton, 1965; Gazzaniga and Sperry, 1967). Differences in manual preference and skill are easy to show in normal subjects, although their interpretation raises problems, particularly with regard to alleged "left-handers." But differences in listening to speech have proved more elusive and difficult to demonstrate in healthy subjects. For example, Palmer (1964) found no difference in thresholds for words, between the right and left ears (although the discrepancy between the ears was slightly greater for those who did better with the right ear). The difficulty in showing differences is not surprising when one considers that stimuli arriving at either ear are relayed to both hemispheres. Most of the striking asymmetries in speech function are shown only when the speech areas of the hemisphere dominant for language are impaired in some way, whether by brain lesions or by the administration of intra-arterial sodium amytal (Penfield and Roberts, 1959), or, more recently, when the cerebral hemispheres have been separated surgically (Gazzaniga *et al.*, 1962, 1967). In these patients it has been possible to test directly the extent to which speech and language are bilaterally represented. Gazzaniga and Sperry (1967) showed that, although the expression of language is confined exclusively to the dominant hemisphere, some comprehension of both written and spoken words and phrases is possible with the minor hemisphere. Two patients were able, for example, to pick out with the left hand an object whose name or description they had just heard.

Thus, in comparing the reception of speech with the left and right ears, two factors must be considered: (1) the degree of lateralization of speech perception; (2) the degree of asymmetry in the representation of the two ears at the ipsilateral and contralateral hemispheres. These two factors are difficult to separate in the performance of normal subjects. Even patients with left hemisphere lesions, provided they are not aphasic, do no worse in the reception of normal speech than patients with lesions in the right hemisphere. However, distorted speech is worse perceived when arriving on the ear contralateral to the lesion (Calearo and Antonelli, 1963). This suggests that the pathway from the contralateral ear (whether left or right ear), may be more important in speech perception than the ipsilateral pathway. Rosenzweig (1951) found that, in cats, auditory inputs affect the two hemispheres unequally, the amplitude of the ipsilateral response being appreciably less than the contralateral one; latency and duration were equal.

More recently, Kimura (1961), Milner (1962), Dirks (1964) and Satz *et al.* (1965), have shown that asymmetries in the performance of normal subjects can be revealed by overloading their limited capacity. Although a message to either ear is handled equally well when given alone, if two messages are presented together, response to the message reaching the ear on the same side as the hemisphere dominant for speech is usually less efficient. To what extent is this asymmetry in speech perception attributable to an asymmetry in the distribution of attention, the result of the incoming messages exceeding the subject's limited capacity?

In a recent experiment on selective attention to competing speech messages (Treisman and Geffen, 1967) we tried to discover what stage in the sequence from receptor to response determined the limit to our capacity for transmitting information. We presented two different speech messages simultaneously, one to each ear, and required two different responses: the subject "shadowed" one message and also monitored both messages for the occurrence of certain specified target words. If one of these target words occurred in either message, the subject was to tap with a ruler. Our subjects tapped to many more targets in the shadowed message, 86·5 per cent. compared to 8·1 per cent. Since the target words and the response load for the two messages were equated in every way, we concluded that selective attention primarily restricts perception, reducing the extent to which the secondary, non-shadowed message can be analysed. We also found some degree of response interference between tapping to and repeating words in the primary message. Subjects occasionally missed one of the two responses while making the other.

This technique can also be used to investigate the relative importance of perceptual attention and of response competition in determining the left–right asymmetry in the reception of and response to speech. In a control condition in our earlier experiment, we compared performance with left and right ears. Subjects always tapped with the left hand to left ear targets and the right hand to right ear targets. There was no difference between the ears in the number of primary targets detected, but more secondary targets were heard on the left ear. More words were correctly repeated on the right ear. We felt these differences were worth exploring further. The two speech messages are given simultaneously to the right and left ears. Both are presumably analysed chiefly by the left hemisphere in almost all right handed subjects (Penfield and Roberts, 1959). The verbal response of repeating back one message is controlled exclusively by the left hemisphere, while the manual response of tapping to target words can be made by either right or left hand, and so is controlled by either left or right motor cortex. Which of these factors or combinations of factors would be likely to reveal asymmetries in performance?

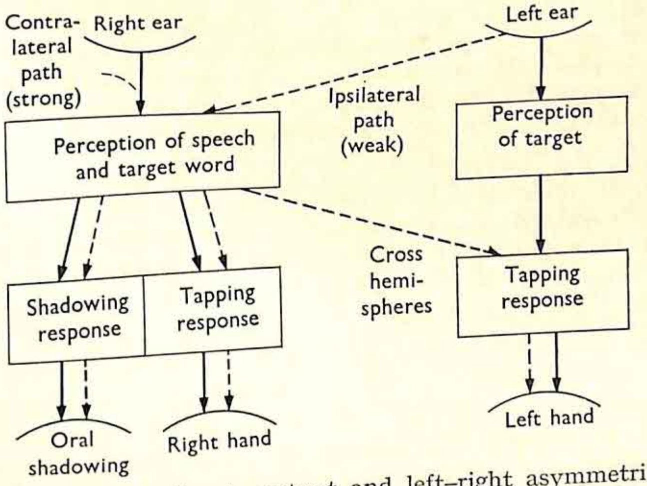
Firstly, do the left and right ear inputs to the left hemisphere speech area differ qualitatively, in some way determined by differences in the anatomical pathways, or only quantitatively? If the asymmetry were quantitative only, it might be eliminated completely by a counter-balancing bias in attention, and show up most strongly when the subject was biased away from the left ear both by the task and by the natural right ear dominance.

Secondly, does the asymmetry vary with the nature of the response required? Both the verbal response of shadowing and the manual response with the right hand are organized in the same hemisphere in which the speech is analysed. Is the input asymmetry between right and left ears more apparent when the response is made with the left hand? Or, since the right ear dominance applies only to speech messages, is the asymmetry revealed only by verbal responses which require some linguistic analysis of the message (by shadowing, but not by tapping). Possibly subjects might even be able to monitor for one specified, simple speech sound equally well with either hemisphere, so that the tapping response might be affected only by hand and not by ear dominance. If so, the verbal context should have less effect on performance with the left ear, since the context would be analysed mainly by the left hemisphere.

Thirdly, how does the asymmetry affect response competition between the two responses to primary target words? Is this determined anatomically? For example, is the competition greater when both responses are organized by the same hemisphere? Or does it depend on functional capacity, so that interference with the repeating response is greater when the left hand is tapping because this is less skilled and so requires more of the response capacity available? Finally is there any demonstrable stimulus-response compatibility, such that the right hand taps more easily to right ear inputs and the left to left ear inputs? The latter would be possible if, for example, the monitoring task could be carried out in the right as well as the left hemisphere. Then perception of the target and response to it could perhaps be organized in parallel for the two messages by the two hemispheres.

Figure 1 may help to clarify some of the problems we have raised. Our experiment certainly does not give final answers to these questions, but it gives some indications of which factors are likely to be the most important. It also helps to confirm or clarify some points in our earlier experiment.

FIGURE 1



Possible relations between input, output and left–right asymmetries. (Potentially weaker paths are shown by dotted lines. These asymmetries could be cancelled out or increased when different biases are introduced by attention.)

METHOD

Subjects. There were 12 right-handed subjects aged between 19 and 31 years. All were self-classified right-handers. They were asked whether they preferred their left hand for any kind of task, and only one mentioned an instance—(she used her left hand for back-hands in tennis). They were volunteers, all Oxford undergraduates, and were paid four shillings an hour.

Stimulus material. Sixty-four 150-word extracts from Conrad's novel *Lord Jim* were prepared as follows: each passage had one target word inserted at three different points. Four different target words were used: "night, trees, hot, and tired," and each was used in eight pairs of passages. In half the passages, the target word did not fit into the context of the passage, and in the other half the context was adapted, if necessary, so that the target words fitted coherently into the structure and sense of the verbal context. These passages were recorded in pairs by a woman speaker on a Ferrograph two-channel tape-recorder, so that a passage on the upper track was paired with a passage on the lower track containing the same target word, both being either in or out of context. All target words on either ear in each pair of passages were separated by at least eight words and none were given in the first or last 10 words in the passage. All passages lasted about 1 min. They were spoken in a monotone and with pauses minimized.

Procedure. The experiment took two, 1-hr. sessions which were held on separate days. The first session was preceded by half an hour of practice at shadowing (repeating back one of the two messages as it was presented). Subjects were first asked to equate the loudness of the dichotic messages, listening to both at once. They then practiced shadowing an equal number of passages on their left and right ears, first without and then with competing messages in the other ear, until they appeared fluent. Finally they shadowed four further passages (one for each condition described below) with the word "tap" inserted out of context three times in both the attended and unattended passages. They were instructed to tap with the ruler every time they heard the word "tap," whether they heard it on the shadowed ear or on the other ear. The primary task was to shadow the selected message and they were to attend primarily to this. They were warned not to shift their attention to the other ear, since this would lead at once to shadowing errors. If any target words "came through" from the non-shadowed message despite the fact that their attention was centred elsewhere, they were to tap. The few cases where the subject had clearly shifted before the target word and was attending to the non-shadowed message were discarded. (The criteria used in discarding these results were that the subjects both reported switching, and also missed several words in the shadowing response just preceding the target word. This topic is discussed in detail in the earlier paper, Treisman and Geffen, 1967.) Two subjects who never tapped to the non-shadowed ear during the four practice passages were discarded. Six subjects who tapped only once to the non-shadowed ear during practice had the volume of the non-shadowed message increased by 2 dB. throughout the experiment proper, since we were interested now in comparing left and right ears rather than determining the absolute effects of attention and inattention.

The experiment was then begun. The four main conditions were:

(1) shadow right ear, tap with right hand to target words in both messages;
(2) shadow right ear, tap with left hand;
(3) shadow left ear, tap with right hand;
(4) shadow left ear, tap with left hand.

There were 16 pairs of passages in each session, two in each condition both in and out of context, and subjects rested for 5 min. after the first eight passages were completed. The order of the passages in the first session was counterbalanced with the opposite hand in the second session for each subject. For example, if "Hot" in context was shadowed last in Session 1 in condition "shadow right ear, tap with right hand," "Hot" in context would be shadowed first in Session 2 in condition "shadow right ear, tap with left hand." All subjects were run in all conditions.

RESULTS

Table I gives the main findings. It shows for each ear–hand condition the percentages of correct tapping responses and shadowing responses to target words and to "other words" (all words except the target words, the three preceding them and the five succeeding them). There was considerable variation in the level of performance between different subjects, as shown by the standard deviations. The results were therefore analysed by comparing individual subject's performances in the different conditions. A number of analyses of variance (all within subjects) were carried out on scores transformed to radians. The results can be summarized as follows.

TABLE I

PERCENTAGE WORDS CORRECT IN TAPPING AND SHADOWING RESPONSES TO THE TWO MESSAGES

	Shadow target word				*Shadow other words*		*Tap to target in shadowed message*				*Tap to target in other message*	
	In context		Out of context				In context		Out of context			
*Condition**	*Mean*	*s.d.*†	*Mean*	*s.d.*	*Mean*	*s.d.*	*Mean*	*s.d.*	*Mean*	*s.d.*	*Mean*	*s.d.*
R_ER_H	95·8	4·5	86·1	15·1	89·7	4·6	95·1	5·3	88·9	11·5	25·0	13·7
R_EL_H	95·1	6·0	86·1	11·9	89·5	4·4	92·4	8·1	88·2	9·0	26·4	18·5
L_ER_H	91·7	8·3	81·9	13·6	82·7	8·5	91·0	9·5	88·2	12·4	20·5	17·5
L_EL_H	95·1	5·1	77·1	21·1	83·8	8·6	93·0	7·1	88·2	12·2	20·1	14·7

* R_ER_H = Right Ear Right Hand; R_EL_H = Right Ear Left Hand, etc. The ear shown here is always the ear receiving the target words in the passage to which the results apply.

† The standard deviations are derived from the transformed (arc-sine) data. They reflect the variations *between* subjects, while the Analysis of Variance (described in the results) was based entirely on *within* subject comparisons of the different conditions.

(*a*) *Shadowing target words*. Subjects correctly shadowed rather more targets in the right than the left ear, but the difference failed to reach significance because of the high inter-subject variation (F ratio = 2·37, $d.f.$ 1, 11, $p < 0{\cdot}2$). As in our previous experiment, target words which fitted into the verbal context were more often shadowed correctly than those inserted at random points (F ratio = 15·0, $d.f.$ 1, 11, $p < 0{\cdot}01$). However, the interaction between ear and context was not significant.

(*b*) *Shadowing other words*. Here the difference in efficiency with left and right ear messages did reach significance (F ratio = 7·6 $d.f.$ 1, 11, $p < 0{\cdot}025$). Subjects were more efficient at shadowing the right ear.

(*c*) *Tapping to target words in the shadowed message* showed no significant effect either of ear or of hand or any interaction; verbal context was again significant (F ratio = 6·0, $d.f.$ 1, 11, $p < 0{\cdot}025$), but did not interact with ear.

(*d*) *Tapping to target words in the non-shadowed message*. Subjects were more likely to tap when the targets were in the right ear and the subject was shadowing the left ear (F ratio = 7·27 $d.f.$ 1, 11, $p < 0{\cdot}025$). There was no significant effect of hand used, and no effect at all of verbal context (F ratio = 0·1). The difference in taps to shadowed and non-shadowed message was again highly significant, as in the previous experiment (F ratio = 283·7, $d.f.$ 1, 11, $p < 0{\cdot}001$).

(*e*) *Response competition*. It is possible also to examine the effects of response interference, given that the target was perceived correctly. Taking only those target words in the shadowed message which did receive a tapping response, a smaller proportion were also correctly shadowed when either the left ear or the left hand were used compared to the right ear or the right hand; (the means for left versus right ear were 93·1 and 96·4, F ratio 5·38, $d.f.$ 1, 11, $p < 0{\cdot}05$, and the means for left versus right hand were 94·4 and 95·7, F ratio = 8·93, $d.f.$ 1, 11, $p < 0{\cdot}025$). Taking only those target words which were correctly shadowed, neither ear nor hand affected the proportion which also received a tapping response. The asymmetry in response interference shows only with the verbal response.

To get some further indication of response intereference, we looked at the effect of tapping on the shadowing response to words around the target as well as on the target itself. Table II shows the mean per cent. of words in the range three before to five after the target which were omitted or wrongly repeated (*a*) when the subject tapped and (*b*) when he failed to tap.

TABLE II

PERCENTAGE ERRORS AND OMISSIONS IN SHADOWING WORDS IN THE RANGE 3 BEFORE TO 5 AFTER TARGET WORD

	Shadowed passage when subject tapped		*Other passage when subject tapped*		*Other passage when subject missed tapping*	
*Condition**	*Mean*	*s.d.*	*Mean*	*s.d.*	*Mean*	*s.d.*
R_ER_H	14·3	6·6	39·2	18·7	13·6	7·7
R_EL_H	16·3	8·3	40·4	23·9	13·2	5·5
L_ER_H	20·1	9·2	31·3	13·2	8·3	5·3
L_EL_H	23·3	12·0	32·1	19·8	8·1	7·0

* The ear shown here is again the ear which received the relevant target words, so that for the "Other Passage" in conditions L_ER_H and L_EL_H, the subject was *shadowing* the right ear and in conditions R_ER_H and R_EL_H he was shadowing the left ear.

(1) *Interference from targets in the shadowed message.*

The analysis of variance for the shadowed message showed significant differences in interference with both ear (F ratio = 5·74, *d.f.* 1, 11, $p < 0·05$) and hand (F ratio = 5·65, *d.f.* 1, 11, $p < 0·05$) when the subject tapped. The efficiency of shadowing these words around the target when the subject tapped was significantly less than his efficiency in shadowing the other words (F ratio 31·7, *d.f.* 1, 11, $p < 0·001$) and, not surprisingly, there was a significant interaction between hand and target-versus-other words. However there was no interaction between ear and target-versus-other words, so there is no evidence here that the extra response of tapping increased the effective asymmetry between the ears on the verbal response of shadowing. The effect of verbal context was not significant.

(2) *Interference from targets in non-shadowed passage when subjects tapped.*

The errors and omissions were higher when the subject shadowed the left ear and tapped to targets in the right ear than the reverse, but the difference here failed to reach significance, probably because this occurred on a small proportion of occasions and so the random variation in general was high. Note that the extent of interference with shadowing varies with the ear receiving the shadowed message rather than with the ear receiving the target words.

There was a significant increase in shadowing errors and omissions for the words around the target compared to other words (F ratio = 95·6, *d.f.* 1, 11, $p < 0·001$). There was no effect of verbal context. The difference in interference when tapping to shadowed and non-shadowed messages was highly significant (F ratio = 613·7, *d.f.* 1, 11, $p < 0·001$).

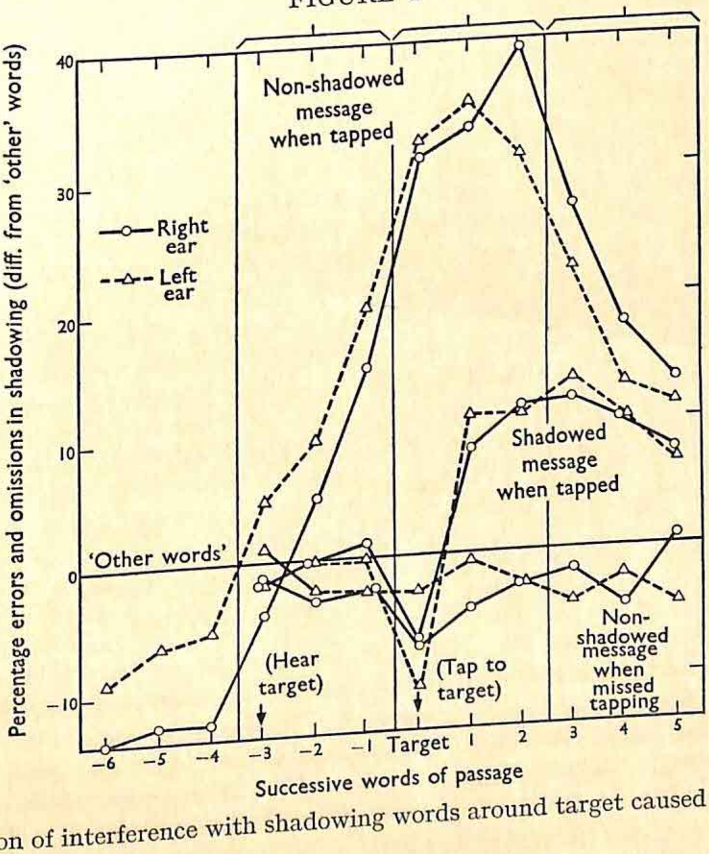
(3) *Interference from non-shadowed targets when the subject missed tapping.*

Here the effect of shadowed ear was again significant (F ratio = 11·97, *d.f.* 1, 11, $p < 0·01$.) Shadowing efficiency to words around the target was significantly greater than to other words, probably because the so-called "other words" included some interference effects from tapping responses to detected target words. The shadowing responses are not back to normal five words after the target; we unfortunately included too few words in our interference range. Certainly there is no evidence of any increase in interference from targets to which the subject made no tapping response. This confirms our previous findings, which we took as evidence that the verbal content of the non-shadowed message is not normally perceived, so that targets which are not tapped to are simply not identified.

Figure 2 represents these results graphically and shows how the interference is distributed over the context of the target words. The errors and omissions have been subtracted from those for shadowing other words in the corresponding hand-ear condition, so the graphs represent deviations from this basal rate of shadowing efficiency. We wondered whether either ear or hand or verbal context would affect the spread of interference or the point at which it peaked in the range of words around the target, but we found no significant interactions with any of these variables.

One point of interest is that the interference for the primary message starts only after the target word has been repeated, while for the secondary message it starts immediately the target word is heard. (In the previous experiment we showed that the lag for both shadowing and tapping responses was about three words.) This suggests that perceiving the secondary targets interfered with repeating back the two or three words which had just been heard, while this was not the case for the primary targets. The difference between secondary and primary messages must be due to having two different stimuli rather than one for the repeating and the tapping responses, since the two tapping responses do not differ in complexity or information load for the primary and secondary message.

FIGURE 2



Distribution of interference with shadowing words around target caused by tapping to target.

The final question which interested us was the relation between ear asymmetry and hand dominance within individuals, as revealed by our experiment. So we carried out a number of product-moment correlations on the differences between left and right ear efficiency and left and right hand efficiency over individual subjects (scores transformed to radians). The results are shown in Table III. One surprising point is how low most of the correlations are. Only one of the possible correlations between asymmetry of ear and of hand reaches statistical significance: the more likely the subject was to tap to a right *ear* primary target, the less interference tapping with his right *hand* would cause. None of the other response measures showed any correlation between right ear and right hand dominance over the individual subjects. Secondly, the inter-correlations between different measures of either ear dominance or hand dominance based on the verbal response of shadowing and the manual response of tapping were also low on the whole, unless the two responses were made to the same word (i.e. the primary target word). The only one which reached significance was the correlation between tapping to primary targets on the right ear and shadowing other words on the right ear. Asymmetries in tapping to secondary targets did not correlate with any other response measure. The only response asymmetry which correlated across primary and secondary message (i.e. despite the disparity introduced by attention) was the negative correlation between the interference from tapping to primary targets and tapping to secondary targets. The less the interference from a right ear target in the primary message (when shadowing the right ear) the more vulnerable to interference the subject was to a right ear target in the secondary message (when he was shadowing the left ear). The effect is probably indirect and due more to which ear the subject was shadowing than to which ear the target word came on. The highly significant correlations are all on closely related or similar response measures such as tapping to and shadowing the same words, shadowing the target and shadowing other words, vulnerability to interference of the shadowing response to words around the target and shadowing efficiency to other words.

TABLE III

Correlations Between Ear and Hand Asymmetry on Various Response Measures

		Primary taps		*Secondary taps*		*Shadow target*		*Shadow other words*		*Primary interference*		*Secondary interference*	
		Ear	*Hand*	*Ear*	*Hand*	*Ear*	*Hand*	*Ear*	*Hand*	*Ear*	*Hand*	*Ear*	*Hand*
Primary taps	Ear												
	Hand	0·04											
Secondary taps	Ear	0·09	0·08										
	Hand	−0·47	−0·39	−0·48									
Shadow target	Ear	0·81‡	0·08	0·24	−0·53								
	Hand	0·03	0·66*	0·16	−0·26	0·20							
Shadow other words	Ear	0·65*	0·16	0·04	−0·21	0·72‡	0·05						
	Hand	0·04	0·44	−0·40	0·09	−0·09	0·36	−0·03					
Primary interference	Ear	−0·54	−0·39	0·07	0·21	−0·62*	−0·19	−0·92‡	−0·11				
	Hand	−0·68†	−0·32	0·04	0·00	−0·31	0·33	−0·28	−0·14	0·06			
Secondary interference	Ear	0·44	−0·37	0·27	−0·10	0·52	0·22	0·76‡	0·16	−0·64*	−0·30		
	Hand	−0·31	0·00	−0·02	0·21	−0·12	0·31	−0·09	−0·51	−0·01	0·47	−0·07	

Level of significance: * $p < 0·025$, † $p < 0·01$, ‡ $p < 0·005$ or higher.

DISCUSSION

This experiment has again confirmed some degree of asymmetry between right and left ear efficiency in speech perception. Most of the earlier experiments (Kimura, 1961; Bryden, 1963; Satz *et al.*, 1965) used Broadbent's dichotic digit recall task (Broadbent, 1954) to reveal the left–right differences, and found it mainly, though not exclusively, in the second half span, which is the most memory-dependent part of the task. Our experiment shows that the differences may also emerge with more immediate responses. The result with tapping to secondary, non-shadowed targets reverses our previous finding: here the right ear targets were detected more often, while in our earlier experiment it was the left ear targets. The present experiment is probably more reliable since it is based on 12 subjects with 16 passages per ear and a mean of more than 20 per cent. tapping responses, compared to 10 subjects, seven passages and about 10 per cent. tapping responses. Nine of the present 12 subjects show the right ear superiority, two show no difference and only one the reverse. Either the earlier finding was due to sampling error, or alternatively it might have been due to the fact that subjects used separate hands to tap to primary and secondary targets in the same pair of passages, the left hand for left ear targets and the right hand for right ear targets. We will discuss this possibility later.

What light do the present results throw on the problems raised in the Introduction? We consider first the asymmetry between right and left ears. To what extent is the right ear dominance simply a quantitative one, due perhaps to an unequal distribution of attention? Our results confirm that the asymmetry can disappear completely, even with competing simultaneous messages, provided that the subject's attention is sufficiently focused on the task. In the primary or shadowed message, left ear targets are tapped to as often as right ear targets and are repeated only slightly less well. The significant decrements can both be attributed to lessened degrees of attention: (1) when the subject is tapping to targets in the secondary, non-shadowed message; (2) when he shadows the "other," non-target words for which is he not specifically set to listen. Notice that the tasks are identical to those showing no asymmetry: only the subject's direction or degree of attention is different. This quantitative asymmetry could be due either to a weaker effective signal-to-noise ratio in messages arriving at the speech areas from the left ear, or to an unequal division of perceptual capacity in the speech mechanisms between right and left ear inputs. Several subjects commented that the left ear became markedly easier to attend to in the second session, after some degree of practice. (We did not compare the sessions quantitatively, since the different conditions were not completely counterbalanced within sessions.) If the decrease in asymmetry were genuine, it would perhaps support the second alternative, since subjects might learn to share the available capacity for speech recognition more efficiently with the left ear, when this carries the more important information.

Secondly, we asked whether the degree of asymmetry between the ears varied with the nature of the response required, whether it would be less apparent with manual responses than with verbal measures of performance requiring linguistic analysis of the stimuli. (Other dichotic experiments have all used verbal responses.) Our results suggest that this is the case. The shadowing response shows the right ear dominance more clearly than the tapping response, particularly when the two responses are competing. Although targets in the shadowed message are tapped to equally well in either ear, the subjects are more likely to miss repeating back these same targets or the words around them when they are on the left ear. However, our results do not make it clear whether this verbal-manual response difference is a separate factor, or whether it too is a result of differences in the distribution of attention. Perhaps our subjects spontaneously put more effort into the tapping than the shadowing task.

The absence of any interaction in tapping response scores between verbal context and ear is rather against the idea that left ear targets are monitored separately in the right hemisphere, since if this were the case, one would expect less effect of verbal context in tapping to left ear targets. In fact there is some suggestion (not statistically significant) that the left ear suffers *more* than the right in the shadowing task when the targets in the primary message are out of context. This difference is in the opposite direction from that predicted if the right hemisphere were used in perceiving the target words. One might also expect better performance for left ear–left hand combinations than left ear–right hand. The difference in the present experiment is minimal.

However, we have to explain the puzzling reversal of dominance for tapping to secondary targets in the earlier experiment. Here the left ear–left hand combination was better than the right ear–right hand combination (12·4 per cent. correct compared to 5·2 per cent.). Target words, subjects and relative intensities of the messages were all different in the two experiments, so we cannot compare absolute values, but the reversal of dominance needs explaining. One possible explanation we tentatively suggest is that the response to be made may influence which hemisphere analyses a particular target stimulus. If the left hand is to make the response, this might increase the likelihood of the right hemisphere being used to monitor the left ear message. We know of no experimental evidence so far either supporting or opposing this view, but it could make sense of the discrepancy between our two results as follows: in the present experiment, the same hand is used in tapping to both primary and secondary targets. On our hypothesis this would make it more likely that both would be analysed by the same hemisphere—the left, speech dominant hemisphere. When the left hand is used for both primary and secondary targets, the right ear primary targets would provide strong competition for the motor areas controlling the left hand, and the left ear secondary targets would suffer.

In the earlier experiment, however, where the left ear secondary targets could monopolize the left hand response systems, these left ear targets might have been analysed separately in the right, non-speech hemisphere, and so escaped much of the competition from the right ear primary message. This would not apply when either (*a*) the *primary* targets were in the left ear (since these also had to be shadowed and this more complex verbal response would require the speech areas of the left hemisphere), or (*b*) the secondary targets were in the *right* ear. In this case, their stronger representation in the left hemisphere and the fact that the response was made with the right hand would encourage analysis by the left hemisphere. Since subjects also had to use this hemisphere for the complex verbal response of shadowing the primary message from the left ear, the competition would be greater and the secondary targets would suffer.

This brings us to the general question of hand dominance and its effect on the results. The only measure in the present experiment on which a difference between the hands appears is the degree of interference with shadowing caused by tapping to primary target words. The right hand dominance shows not in the tapping task itself, but in its less disruptive effect on the shadowing task. Although the response capacity is sufficient to allow equal efficiency in tapping with left and right hands, use of the left hand leaves less spare capacity for the simultaneous verbal response, and this succumbs to the competition. There were no significant interactions at all between ear and hand. This version of stimulus-response compability evidently does not increase the capacity available when the same hand is used in response to both ears. Nor does the use of the same hemisphere for both responses to primary targets (shadow verbally and tap with right hand) impair performance at all. If anything, it facilitates it slightly: the right hand conditions gave slightly fewer errors on most measures. This is more compatible with a functional limit to capacity than with a localized anatomical one.

The absence of correlations within individual subjects between asymmetries of ears and of hands suggests that the degree of asymmetry between left and right ear representation in contralateral and ipsilateral hemispheres is rather loosely related to the degree of right hand dominance.

Finally a few general points may be worth noting in relation to our earlier paper. Firstly, in our earlier experiment we had found a slight but significant effect of verbal context in the secondary, unattended message. This raised problems for our account of attention in terms of a perceptual filter, since subjects should be unable to analyse the verbal content of the message they were not shadowing. We argued then that the result might be an artefact, and the present experiment supports this conclusion, since it shows no effect at all of verbal context in the secondary message.

Secondly, we did not, in the earlier experiment, analyse the way in which the interference caused by tapping was distributed over the words around the target. Although Figure 2 reveals no significant left–right asymmetries, it has some more general interest, since it shows a clear difference in the distribution as well as the extent of the interference for shadowed and for non-shadowed message. One can divide the range of words into three sections. Given that subjects lag an average of three words behind the stimulus both in shadowing and in tapping, words three to one before the target show the interference caused by perception of the target on memory of or response to the three words heard previously, before the target occurred. Perception of targets in the shadowed message caused no interference at all; subjects' shadowing was disrupted only after the tapping response was made and the target word itself was actually better shadowed than the other words when it fitted into the context. In the non-shadowed message, on the other hand, when the target word is heard, the resultant shift of attention interferes with response to the two words which occurred just before it, perhaps wiping out the memory trace before perception is complete. The target word and the two words following it show, for both ears, the effects of response competition: tapping to the target interferes with speaking. For the non-shadowed ear, these words also show the effect of perceptual competition. The considerable number of extra errors and omissions caused by non-shadowed compared to shadowed targets show the extent to which *hearing* the target prevents the subject from *hearing* the competing words in the primary message. Finally the words three to five after the target show, for both ears, how the tapping response interferes with the perception and memory of subsequent words in the shadowed message. Here the results for the two ears come together, as one would expect, since the tapping response itself is the same for both ears. The errors and omissions when subjects missed tapping to secondary targets show no evidence of interference at all and remain more or less constant over the range of words around the target, as one would expect if these targets were simply not perceived.

This work was written up whilst the senior author was at Bell Telephone Laboratories, New Jersey.

We should like to thank the Medical Research Council for supporting the research and P. S. Vivekananthan, who helped in analysing the results.

REFERENCES

BENTON, A. L. (1965). The problem of cerebral dominance. *Canad. Psychologist*, **6**, 332–48.

BROADBENT, D. E. (1954). The role of auditory localization and attention in memory span. *J. exp. Psychol.*, **47**, 191–6.

BRYDEN, M. P. (1963). Ear preference in auditory perception. *J. exp. Psychol.*, **65**, 103–5.
CALEARO, C., and ANTONELLI, A. R. (1963). "Cortical" hearing tests and cerebral dominance. *Acta otalaryng.*, **56**, 17–26.
DIRKS, D. (1965). Perception of dichotic and monaural verbal material and cerebral dominance for speech. *Acta otolaryng.*, **58**, 73–80.
GAZZANIGA, M. S., BOGEN, J. E., and SPERRY, R. W. (1962). Some functional effects of sectioning the cerebral commissures in man. *Proc. nat. Acad. Sci.*, **48**, 1765–9.
GAZZANIGA, M. S., and SPERRY, R. W. (1967). Language after section of the cerebral commissures. *Brain*, **90**, 131–48.
KIMURA, D. (1961). Cerebral dominance and the perception of verbal stimuli. *Canad. J. Psychol.*, **15**, 166–71.
MILNER, B. (1962). Laterality effects in audition. In MOUNTCASTLE, V. B. (Ed.) *Interhemispheric Relations and Cerebral Dominance*. Baltimore: Johns Hopkins University Press, 177–95.
PALMER, R. D. (1964). Cerebral dominance and auditory asymmetry. *J. Psychol.*, **58**, 157–67.
PENFIELD, W. P., and ROBERTS, L. R. (1959). *Speech and Brain Mechanisms*. London: Oxford University Press.
ROSENZWEIG, M. R. (1951). Representations of the two ears at the auditory cortex. *Amer. J. Physiol.*, **167**, 147–58.
SATZ, P., ACHENBACH, U., PATTISHALL, E., and FENNELL, E. (1965). Order of report, ear asymmetry and handedness in dichotic listening. *Cortex*, **1**, 377–96.
TREISMAN, A. M., and GEFFEN, G. (1967). Selective attention: perception or response? *Quart. J. exp. Psychol.*, **19**, 1–17.
ZANGWILL, O. L. (1960). *Cerebral Dominance and its Relation to Psychological Function*. Edinburgh: Oliver and Boyd.

Manuscript received 11th October, 1967.

ONE-TRIAL LEARNING WITH CONTROL OF ITEM-DIFFICULTY

BY

SALVATORE A. PIZZURRO and JOHN VAN LAER

From PRADCO, New York, and Hunter College, City University of New York

Ratings of difficulty were obtained for each of 2,500 letter-number pairs. The ratings were used to construct three pools of items: Homogeneous Easy, Homogeneous Difficult, and Heterogeneous. Rock's (1957) experiment was conducted thrice, once with each pool of rated items. Each replication used 30 subjects. The manipulation of rated item-difficulty permitted a direct attack on the widely discussed problem of item-selection in Rock's procedure. Notwithstanding the control of item-selection, the performance of subjects in the drop-out condition did not differ significantly from the performance of those who learned in the ordinary way, with repetition. The geometric distribution provided a good fit to the obtained distributions of waiting times to the first correct response to an item. This finding accords well with an all-or-none view of associative learning.

INTRODUCTION

Rock's (1957) paper, setting forth an all-or-none view of the formation of associations in paired-associate learning, released a spate of research in that area. Most subsequent work fits into one of two broad categories. There are those (Postman, 1962; Underwood and Keppel, 1962; Underwood, Rehula, and Keppel, 1962) whose research has been in part motivated by the valid critical objection that the drop-out condition (DOC) employed by Rock and his co-workers (Rock, 1957; Rock and Heimer, 1959; Rock and Steinfeld, 1963) is biased by the elimination of difficult items and the substitution of easier items from the replacement pool. On the other hand, proponents of the all-or-none view (Estes, 1960; Kintsch, 1963; Kintsch and Morris, 1965; Polson, Restle, and Polson, 1965) have concentrated their efforts on detailed analyses of the statistics of learning in order to demonstrate the goodness of fit of data to quantitative predictions made by some model. The several models all view paired-associate learning as a process that encompasses two or more discrete states that an item may or may not enter on any trial.

This study combines elements of both approaches. Underwood, Rehula and Keppel (1962) have shown that ratings of ease of learning have remarkable efficacy in predicting subsequent performance in paired-associate learning. In this experiment control of the difficulty of the material was employed to investigate the effect of item-selection in the drop-out condition. The most salient feature of the all-or-none position viz. that the waiting time to the first correct response is theoretically governed by the geometric distribution, is examined.

Suppose that the initial list presented to subjects in the drop-out condition (DOC) contains items of homogeneous difficulty, and is identical to the list learned by control-group subjects in the ordinary way, with repetition. Further, suppose that the pool from which substitute items are drawn to replace those missed by subjects in DOC is uniformly as difficult as the initial list. If, under these conditions, the subjects in DOC were found to perform as well as the control subjects, such a finding would blunt the widespread criticism that Rock's technique is marred by insurmountable methodological flaws. That criticism doubtless has some validity, *a priori*. The empirical question, as to whether Rock's results are merely artefacts of item-selection, remains open.

METHOD

Material. Rock (1957) used letter–number pairs in which the stimulus element was either a single letter (*I* excluded) or a repeated letter, and the response element was some integer from 1 to 50. In preliminary work, ratings of difficulty on a 5-point scale, wherein 1 denoted Least Difficult and 5 denoted Most Difficult, were obtained for all 2500 such pairs. Rating procedures and instructions to raters were the same as those used by Underwood, Rehula and Keppel (1962). Six groups of five undergraduates rated sets of 100 pairs. These 600 pairs were arranged randomly into six different lists of 100 pairs and again rated by six groups of five raters. Each of the 600 twice-rated pairs yielded two mean ratings, assigned by two distinct groups of five raters in two different item-contexts. The product-moment correlation between first mean rating and second mean rating was 0·83. On the basis of this correlation it was assumed that the rating procedure had sufficient reliability and the remaining 1900 pairs were rated by only five raters each. Seven groups of five rated 240 pairs each, and the last group rated the remaining 220 pairs.

Mean ratings were used to construct three pools of 30 items. The Easy Homogeneous pool contained pairs that received mean ratings between 1·5 and 2·0; pairs in the Difficult Homogeneous pool had mean ratings between 4·0 and 4·5; the Heterogeneous pool comprised eight pairs that had mean ratings between 1·0 and 1·9; seven between 2·0 and 2·9; seven between 3·0 and 3·9; and eight between 4·0 and 5·0. All pools were constructed with the restriction that no two pairs in a given pool have an element in common on either the stimulus- or the response-side.

The Easy Homogeneous pool contained:

N–4, H–5, P–6, X–7, O–20, C–15, A–35, B–22, Q–19, D–11, R–23, E–50, M–13, F–12, S–21, G–25, T–18, U–16, J–30, V–26, K–40, W–33, L–10, Y–42, Z–34, OO–1, FF–2, AA–8, MM–6, ZZ–3.

The Heterogeneous pool consisted of the following pairs, listed in the order of increasing difficulty ratings:

A–7, B–3, C–9, D–8, E–4, T–5, XX–1, QQ–2; F–14, G–43, H–13, K–12, S–17, R–18, Y–21; M–41, N–32, Q–29, V–48, AA–31, BB–42, CC–36; EE–47, FF–49, KK–34, LL–24, UU–23, RR–26, MM–28, VV–46.

The Difficult Homogeneous pool contained:

X–47, S–14, Z–26, V–48, T–53, ZZ–24, XX–13, RR–18, VV–46, WW–23, AA–49, BB–25, CC–39, DD–28, EE–29, HH–34, GG–21, FF–50, KK–36, LL–32, MM–31, NN–42, OO–38, PP–45, JJ–17, YY–12, SS–19, QQ–41, TT–27, UU–37.

Subjects and procedure. Ninety experimentally naive undergraduate psychology students were assigned at random to experimental and control groups, with the restriction that there were 45 in each group. Experimental and control groups were divided at random into subgroups of 15 subjects each. One control and one experimental subgroup worked with each of the three pools of letter–number pairs. Out of each pool, 15 initial lists of 10 pairs were drawn at random. Each initial list was assigned to one subject in the control group and to one subject in the experimental group. The pairs were typed in the centre of 3 by 5-in. cards. Learning proceeded by the alternate study-and-test-trial method. On test trials, subjects viewed the reverse sides of the cards, which exhibited the stimulus elements only.

The criterion was one errorless repetition. Control subjects retained the initial list throughout training. Experimental subjects underwent DOC, wherein the experimenter replaced a pair missed on any trial by a pair chosen at random from the remaining items in the pool. On study trials, the cards were exposed for 3 sec., with a 5-sec. interval between successive presentations. On test trials, the stimulus element was exposed for 5 sec., and there was a 3-sec. interval between successive items. The interval between the last item of a study trial and the first item of a test trial was 15 sec. A 30-sec. interval followed each test trial. During this interval, the experimenter made substitutions for pairs missed by subjects in DOC, and shuffled the deck to rearrange the serial order of presentation for all subjects. Rock's procedure was followed, except for list-length (10 vs. 12), size of pool (30 vs. 50), instructions, and the manipulation of item-difficulty.

Instructions differed from those used by Rock (1957) in two ways. (*a*) Both experimental and control subjects were instructed to *learn* all the pairs shown on a given trial. (In Rock's experiment, the control group was told to *associate* stimuli and responses—a circumstance that led Reed and Riach (1960) to argue that Rock's result may have been peculiar to his instructions). (*b*) During study trials, the subjects were required to speak the stimuli and responses aloud, and on test trials they spoke the stimuli aloud whether or not they could make the correct responses.

RESULTS

Means and standard errors of trials to criterion are presented, by subgroups, in Table I. Note that means of the control and experimental subgroups whose lists came from a given pool were nearly identical. The only apparent difference was between those subgroups that learned the Easy items. That difference did not approach statistical significance: $t = 1{\cdot}4$, 28 *d.f.*, $p > 0{\cdot}20$.*

TABLE I

MEANS AND STANDARD ERRORS OF TRIALS TO CRITERION

	Experimental		*Control*	
Item pool	*Mean*	SE_m	*Mean*	SE_m
Easy Homogeneous	3·5	0·28	3·0	0·21
Heterogeneous	4·3	0·49	4·2	0·42
Difficult Homogeneous ..	5·2	0·49	5·2	0·38

The validity of the difficulty ratings was examined by correlating the mean ratings with the "selection ratio" (Underwood *et al.*, 1962), i.e. the ratio of the number of times an item appeared in the criterial list of subjects in DOC to the total number of times it might have appeared. For items in the Heterogeneous pool, the product-moment correlation between selection ratio and rated difficulty was $-0{\cdot}68$, which was significantly different from zero at the 0·01 level. The analogous correlations in the two Homogeneous pools were both exactly zero. Moreover, Table I shows that mean trials to criterion for the various subgroups were in the appropriate rank order of rated difficulty. Data of both subgroups that shared items at a given level of difficulty were pooled. Comparisons via Tukey's method showed that the pairwise differences among the pooled mean trials to criterion were all significant at the 0·01 level.

The distribution of the number of trials preceding the first correct response to a given item was examined. Provided the subjects are fairly homogeneous in learning ability, and provided the items are of uniform difficulty, there is some initial probability, P, that the subject gives the correct response after a single exposure to the item. The all-or-none view asserts that repetition, prior to emission of a correct response, leaves P unchanged. This in turn implies that the distribution of the number of trials that precedes the first correct response is the geometric distribution. If the all-or-none view is true, the probability that the first correct response is preceded by exactly k misses is Q^kP, where $Q = (1-P)$.

Analysis of the distribution of proportions, in each subgroup, of initial correct responses that were preceded by a given number of misses yielded six values of $\hat{P}$, the maximum likelihood estimate of P. These estimates generated six theoretical

---

* At the suggestion of a referee, the significance of the difference between mean trials to criterion was subjected to re-analysis via the Mann–Whitney test. The smallest of the three obtained values of U was 88, whereas U must equal at most 72 in order to be statistically significant at the 0·05 level in a one-tailed test.

TABLE II

Observed (*Obs.*) and Theoretical (*Theo.*) Distributions of the Number of Trials Preceding the First Correct Response

Number of trials (k)	*Homogeneous Easy*				*Heterogeneous*				*Homogeneous Difficult*			
	Experimental		*Control*		*Experimental*		*Control*		*Experimental*		*Control*	
	Obs.	*Theo.*	*Obs.*	*Theo.*	*Obs.*	*Theo.*	*Obs.*	*Theo.*	*Obs.*	*Theo.*	*Obs.*	*Theo.*
0	0·467	0·503	0·553	0·601	0·440	0·460	0·453	0·451	0·367	0·362	0·360	0·374
1	0·233	0·250	0·280	0·240	0·280	0·248	0·233	0·248	0·207	0·231	0·207	0·234
2	0·193	0·124	0·127	0·096	0·113	0·134	0·127	0·136	0·140	0·147	0·153	0·147
3	0·073	0·062	0·033	0·038	0·073	0·072	0·080	0·075	0·073	0·094	0·107	0·092
4	0·027	0·031	0·007	0·025	0·060	0·039	0·073	0·041	0·113	0·060	0·100	0·058
5	0·007	0·030	—	—	0·020	0·021	0·020	0·022	0·053	0·038	0·047	0·036
6	—	—	—	—	0·013	0·025	0·013	0·027	0·013	0·024	0·003	0·022
7	—	—	—	—	—	—	—	—	0·027	0·016	0·010	0·014
8	—	—	—	—	—	—	—	—	0·007	0·027	0·013	0·024

Note.—Theoretical values are $\hat{Q}^k\hat{P}$, where $\hat{P}$ is the leading entry in the *Theo.* column.

geometric distributions, one for every subgroup. Table II exhibits the actual data, as well as the fitted geometric distributions. Inspection of Table II shows that, as Polson, Restle and Polson (1965) reported, the geometric distribution fits the data adequately. Table III contains the results of chi-square tests of goodness of fit of theoretical to obtained values. Only one chi-square statistic, that for the group in DOC that learned Easy items, approached significance at the 0·05 level. The data do not exhibit the anticipated departure from predictions based on the hypothesis of one-trial learning.

TABLE III

Chi-square Tests of Goodness of Fit of the Geometric Distribution to the Observed Distribution of the Number of Trials Preceding the First Correct Response

Subgroup	χ^2	*d.f.*	*p*
Easy Control	5·28	3	>0·10
Heterogeneous Control	5·24	5	>0·25
Difficult Control	9·53	7	>0·10
Easy Experimental	9·49	4	=0·05
Heterogeneous Experimental	3·69	5	>0·50
Difficult Experimental	10·24	7	>0·10

Comparison of the distribution of trials to the first correct response, for DOC and control subgroups, follows. For the Easy items, a chi-square test of homogeneity of control and experimental distributions yielded $\chi^2 = 8{\cdot}70$, 4 *d.f.*, $0{\cdot}05 < p < 0{\cdot}10$. For Heterogeneous items, $\chi^2 = 2{\cdot}02$, 5 *d.f.*, $p > 0{\cdot}50$. For Difficult items, $\chi^2 = 2{\cdot}26$, 6 *d.f.*, $p > 0{\cdot}50$. These sum to yield $\chi^2 = 12{\cdot}98$, 15 *d.f.*, $p > 0{\cdot}50$. The test statistic for DOC vs. control subgroups that learned Easy items approached significance, but when this result was combined with the others, the pooled χ^2 for the entire experiment did not approach statistical significance.

The values of $\hat{P}$ that appear in Table II confirm the predictive validity of the difficulty ratings. $\hat{P}$ is the entry at the top of any column headed "*Theo.*" If the ratings are valid, $\hat{P}$ for Easy items should exceed $\hat{P}$ for Difficult items. Inspection of Table II shows that these results were in the expected direction. An appropriate statistical test for the significance of the differences is not immediately obvious. However, standard methods (Mood and Graybill, 1963, p. 236) give

$$P^2Q/N(1-Q^m)$$

as the asymptotic variance of $\hat{P}$, when the maximal observed run of errors preceding a first success is of length m. Since $\hat{P}$ is "best asymptotically normal" (Mood and Graybill, 1963, p. 174), the ratio of the difference between two independent values of $\hat{P}$ to the estimated standard error of that difference is a test statistic that has (on the null hypothesis) the unit normal (Z) distribution asymptotically. For control subgroups, the difference between $\hat{P}$ for Easy and $\hat{P}$ for Difficult items yielded $Z = 5{\cdot}85$; the Z for experimental subgroups was $3{\cdot}70$. Both comparisons were highly significant.

DISCUSSION

This experiment attempted to determine whether Rock's (1957) result would fail to appear when item-difficulty in the replacement pool was controlled. Experimental control of item-difficulty was achieved by assembling pools of items at different levels of rated difficulty. In the experimental group that learned Heterogeneous items, the correlation between an index of selection and rated difficulty was substantial. In the Easy and Difficult pools, this correlation was zero, reflecting adequate restriction of the range of item-difficulty. All other statistical comparisons among the pools of items were in the appropriate direction. It may seem inconsistent that mean trials to criterion were nearly identical for control and DOC subjects that used the Heterogeneous pool, notwithstanding the prevalence of item-selection by subjects in DOC. However, this result merely confirms the attainment of a uniform distribution of difficulty in the Heterogeneous pool. Uniformity implies that an item randomly picked from the pool is, on the whole, just as likely to be difficult and dropped in turn as it is to be easy and retained. A subject in DOC finishes with an easier list, eventually, after some false moves. Nevertheless, he is on the average no better and no worse off than a control subject who retains an initial randomly selected list throughout learning.

Although item-difficulty was held constant, there were no significant discrepancies between the performance of subjects in DOC and that of control subjects who learned homogeneous materials, in mean trials to criterion or in the number of errors preceding the first correct response to each item. The insensitivity of the data to the gross procedural variation between DOC and control treatment is all the more striking in the light of their reliable and appropriate response to the manipulation of subjective ratings of item-difficulty. The latter is, on the face of it, a rather tenuous independent variable compared to the occurrence or nonoccurrence of repetition.

It is highly unlikely that the failure of control subjects to outstrip DOC subjects in this experiment was attributable to progressive selection of easier items by the DOC subjects. In addition, intralist similarity was minimized in all three pools of items. Hence, a counterargument that performance in DOC was artefactually enhanced by the deletion of items that contributed excessive intralist interference would be groundless. The remaining viable inference is that the subjects in DOC were not handicapped by the absence of repetition of unlearned items.

There was good agreement between the observed distributions of waiting times and the geometric distribution. The fit of the geometric distribution to the waiting-time data signifies that the proportion, Q, of unlearned items that are again missed after an additional repetition remains fixed, even after several unsuccessful repetitions. This comes as no surprise in DOC, where unlearned items are not repeated. In DOC, the fit of the geometric distribution merely confirms that the statistical process of learning is appropriately stationary. In the control condition, however, this result runs counter to intuition and argues strongly in favour of the all-or-none point of view.

This research was supported in part by contract Nonr 477 (34) between the Office of Naval Research and the University of Washington. Reproduction in whole or in part for any purpose of the United States Government is permitted.

The writer's debt to E. H. Galanter for valuable editorial suggestions and material help in preparing this paper for publication is here acknowledged.

REFERENCES

ESTES, W. K. (1960). Learning theory and the new "mental chemistry." *Psychol. Rev.*, **67**, 207–23.

KINTSCH, W. (1963). All-or-none learning and the role of repetition in paired-associate learning. *Science*, **140**, 310–2.

KINTSCH, W., and MORRIS, C. J. (1965). Application of a Markov model to free recall and recognition. *J. exp. Psychol.*, **69**, 200–6.

MOOD, A. M., and GRAYBILL, F. (1963). *Introduction to the Theory of Statistics*. 2nd Ed. New York: McGraw-Hill.

POLSON, M. C., RESTLE, F., and POLSON, P. G. (1965). Association and discrimination in paired-associates learning. *J. exp. Psychol.*, **69**, 47–55.

POSTMAN, L. (1962). Repetition and paired-associative learning. *Amer. J. Psychol.*, **75**, 372–89.

REED, J. C., and RIACH, W. D. (1960). The role of repetition and set in paired-associate learning. *Amer. J. Psychol.*, **73**, 608–11.

ROCK, I. (1957). The role of repetition in associative learning. *Amer. J. Psychol.*, **70**, 186–93.

ROCK, I., and HEIMER, W. (1959). Further evidence of one-trial associative learning. *Amer. J. Psychol.*, **72**, 1–16.

ROCK, I., and STEINFELD, G. (1963). Methodological questions in the study of one-trial learning. *Science*, **140**, 822–4.

UNDERWOOD, B. J., and KEPPEL, G. (1962). One-trial learning? *J. verb. Learning verb. Behav.*, **1**, 1–13.

UNDERWOOD, B. J., REHULA, R., and KEPPEL, G. (1962). Item-selection in paired-associate learning. *Amer. J. Psychol.*, **75**, 353–71.

Manuscript received 17th November, 1967.

PERIODICITY WITHIN REACTION TIME DISTRIBUTIONS AND ELECTROMYOGRAMS

BY

M. RUSSELL HARTER* and C. T. WHITE

From the University of Miami, Coral Gables, Florida and Naval Electronics Laboratory Center, San Diego, California

This study investigates a periodic component in reaction time frequency distributions, that is, a tendency for responses to occur at regular, discrete intervals of time after stimulus presentation. Reaction time frequency distributions were plotted by a Computer of Average Transients and were obtained under stimulus conditions varying in sense modality stimulated (auditory and visual), and the intensity, colour, and duration of stimulation. The results indicated that there was periodicity in reaction time frequency distributions with a modal period of approximately 25 msec. It was found that the periodicty (*a*) was most evident when there was considerable variability in reaction time, and (*b*) tended to attenuate when a large number of reaction times were grouped. Other stimulus conditions appeared to have little effect on the periodicity. A significant correlation was found between the frequency of periodicity in the reaction time distributions and the electromyograms, both having a modal period of 25 msec. It was concluded that the periodicity in reaction time was the result of motor processes.

INTRODUCTION

Latour (1961, 1967) and Latour and Bouman (1961) noted that visual pursuit movement reaction times were not normally distributed but had a periodic nature. That is, movement tended to occur at discrete, periodic points in time after the stimulus was presented. Of particular interest, they found periodicities with frequencies of 100 cps., much faster than the 10 to 20 cps. noted elsewhere (Augenstine, 1955; Eason, Oden, and White, 1967; Harter, 1967; Venables, 1960). The purposes of the present study are (*a*) to demonstrate that manual reaction times to both auditory and visual stimuli have a similar periodic nature; and (*b*) to relate this periodicity to periodicity within electromyograms.

METHOD

Reaction times (RTs) were obtained by having subjects respond to 10-microsec light flashes or auditory clicks (generated by a Grass PS-2 Photo-stimulator) by pushing a microswitch key with their right index finger. All subjects were experienced with the RT task. The time between stimulus presentation and the subject's response was measured by a Hewlett Packard 5223L Electronic Counter and recorded by a Hewlett Packard J74 562A Digital Recorder.

A Computer of Average Transients (CAT) was used to obtain an on-line measure of the reaction time distributions. The stimulus light flash triggered the CAT sweep and the subject's response caused a 10 msec. square pulse to be added into the CAT memory bins (the bins associated with the point in time when the subject responded). Each time-bin stored energy over a 5 msec. interval. Therefore, the 10 msec. pulse, associated with the response, was distributed evenly in two bins or unevenly in three bins, the distribution of energy in the bins giving information as to exactly when the response occurred. When a number of responses or pulses were summed, the voltage level in the time-bins reflected the number of responses occurring at any given point of time after stimulus presentation. A Moseley 135C Autograph (X-Y plotter) was then used to record the distributions on graph paper.

It should be noted that the 10 msec. pulse was selected for a number of reasons: (*a*) it gave the optimal bin-width to pulse-duration ratio; (*b*) it had a rounding effect on the distributions attenuating the activity with a period of less than 10 msec. but having little effect on other activity; and (*c*) it enabled a perceptually meaningful frequency distribution to be plotted when relatively few RTs were combined.

---

* Now at University of North Carolina at Greensboro, Greensboro, North Carolina, 27412.

This technique for plotting frequency distributions created the optimum conditions for observing the periodicity under investigation in the present study. With the usual method for plotting RT data, data are placed into intervals and are assumed to be normally distributed about the mid-points of the intervals. This assumption could not be met in the present study. Furthermore, if the interval had been too large or too much data had been combined, the phenomenon under investigation would have been masked. With the computer technique, there was functionally no interval in the frequency distributions due to the bin-width and RT pulse used as described above and distributions could be plotted after relatively few RTs were combined. Furthermore, the results could be seen on the oscilloscope at all times during data accumulation which enabled the experimenters to observe the development and nature of the RT distributions.

RESULTS

We examined more than 300 RT frequency distributions, containing 25–200 RTs each, which were obtained under stimulus conditions varying in stimulus modality, colour, intensity, and duration. Visual inspection of these distributions revealed that they had a definite periodic component and that stimulus conditions *per se* had little effect on the periodicity, except in relationship to how they influenced RT variability as considered below. Therefore, the various stimulus conditions utilized will not be reviewed in detail here, but will be mentioned briefly below in regard to specific data and problems.

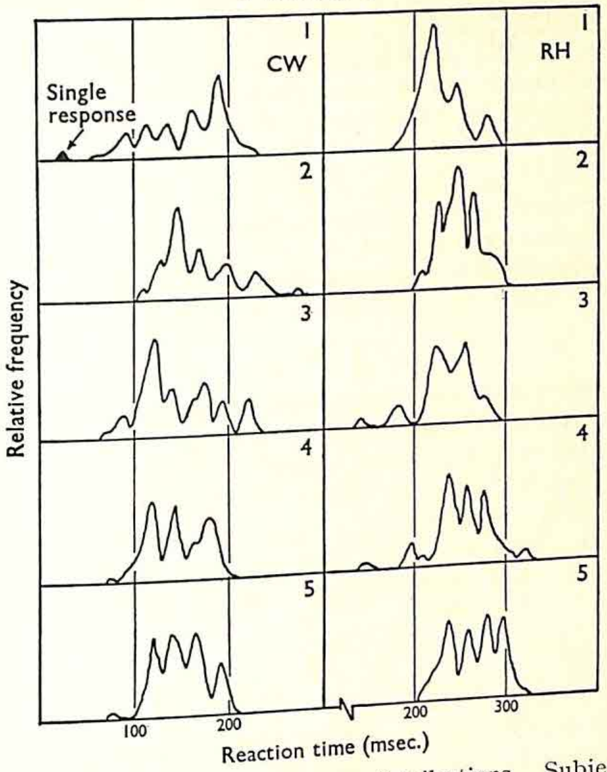
Nature of periodicity

An illustration of the periodicity manifest in the distributions is seen in Figure 1. These distributions were selected from 36 RT distributions obtained by having subjects respond to the offset of flash-trains which contained from 3 to 12 flashes presented at 33 cps. The subjects binocularly fixated a point 2° above the flash source, which subtended a visual angle of 17 min. The flashes (10 microsec. in duration) had a square wave onset, were approximately 2·5 log units above relative threshold, and were surrounded by a homogeneous white field of 18 ft.-L. The flash-trains were presented 2 sec. after the subject pressed a "ready" key with his left index finger.

These data illustrate a number of general findings. First, the RT distributions tended to be multi-modal; responses tended to occur most frequently at a number of discrete points in time after the stimulus was presented. Second, the incidence of the periodicity within each of the 36 distributions varied considerably. And third, the points in time when the subjects responded most and least frequently, that is the peaks and troughs of the distributions, varied from one replication to the next.

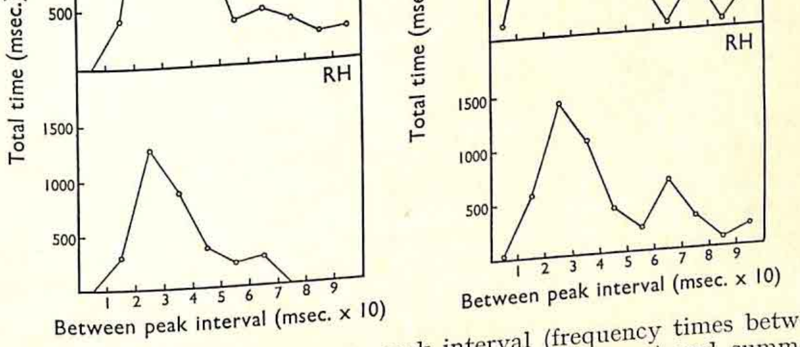
In order to determine the modal period or interval of this activity, frequency distributions of the between-peak intervals in the original 36 RT distributions were made for each subject. The frequency of occurrence of each interval was then weighted by the duration of the interval. (This measure was selected so that the same weight would be given to long between-peak intervals as to short ones. A simple frequency count neglects the fact that more short intervals can occur in a given interval of time than long ones. For example, an RT distribution with a range of 100 msec. may have four 25 msec. between-peak intervals but only two 50 msec. between-peak intervals although both intervals were evident 100 per cent. of the time.) The resulting distributions (see Fig. 2) indicate the total time occupied by each between-peak interval. The modal between-peak interval for both subjects in the RT data was 25 msec. (40 cps.) with a suggestion of a second modal interval of 65 msec. These values are consistent with those reported by Latour (1961, 1967) for visual pursuit movement.

FIGURE 1



Examples of periodicity in RT frequency distributions. Subjects CW and RH responded to the offset of flash-trains of 3–12 flashes presented at 33 cps. The five replicated distributions for each subject are each based on an N of 50. An upward deflection indicates increased frequency.

FIGURE 2

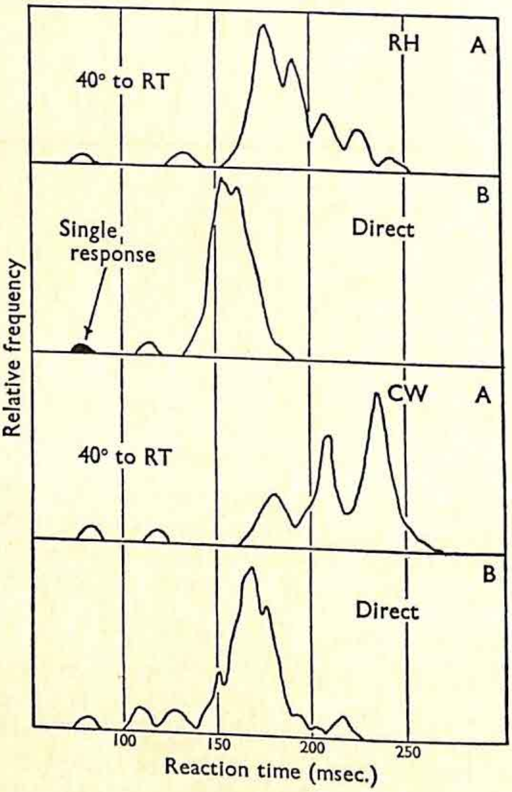


Total time occupied by each between-peak interval (frequency times between-peak interval) in subjects CW's and RH's RT distributions (left figure) and summed EMG records (right figure). The plotted points are mid-points of successive 10 msec. intervals (see text for further explanation).

RT variability and periodicity. It may be noted in Figure 2 that there was a greater incidence of periodicity in subject CW's RT distributions than in subject RH's RT distributions. For example, there was 1,900 msec. of the modal interval in subject CW's records as compared to 1,250 msec. of the modal interval in subject RH's records. This may have been due to the greater variability of subject CW's RTs, which resulted in a larger sample of time, as delineated by the range of each RT frequency distribution, from which to measure periodicity.

The relationship between RT variability and the incidence of periodicity therefore was investigated by manipulating variables known to influence RT variability (stimulus intensity, background intensity, and point of retinal stimulation). The four distributions presented in Figure 3, which were obtained by having subjects respond to single light flashes presented to the temporal and foveal retina, illustrate the relationship found between RT variability and the periodicity. The data from both subjects indicate that the periodicity was more evident in the distributions with greater variability (distributions A) than in those with lesser variability (distributions B). It may be noted that this relationship held regardless of the variables used to manipulate RT variability.

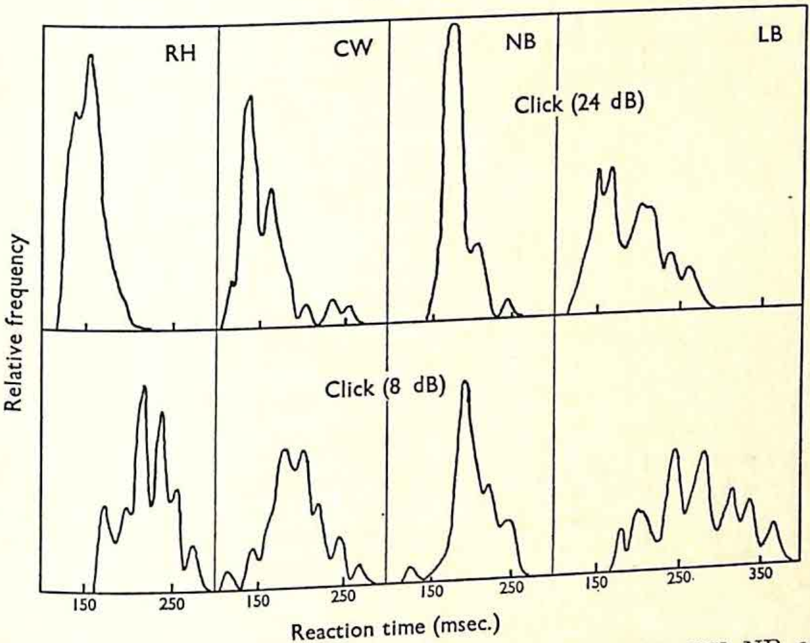
FIGURE 3



Examples of RT frequency distributions given by subjects RH and CW in response to single light flashes presented to the temporal (40° from fovea) (distributions A) and foveal (distributions B) retina. Except for the subjects viewing the single flashes monocularly, these conditions were identical to those described in conjunction with Figure 1. Each distribution is based on an N of 50. An upward deflection indicates increased frequency.

In order to investigate if the periodicity would occur in RT distributions obtained with auditory stimuli, four subjects responded to clicks 8 and 24 dB. above relative threshold. The clicks were generated by a Grass PS2 Photo-stimulator in conjunction with a Hewlett Packard Model 350B attenuator. It was found that the RT distributions obtained to clicks were similar to those obtained to flashes. The eight distributions given in Figure 4 are representative of the auditory RT data collected from the four subjects (RH, CW, NB, LB) under each of the two click loudness conditions. These data illustrate (1) that responses tended to occur at discrete, regular points in time after the stimulus was presented, the successive points being separated by approximately 25 msec., and (2) that greater periodicity was evident in the RT distributions obtained with the soft click (8 dB.), this condition resulting in greater RT variability. Thus as in Figure 3, the periodicity was more apparent with greater variability.

FIGURE 4

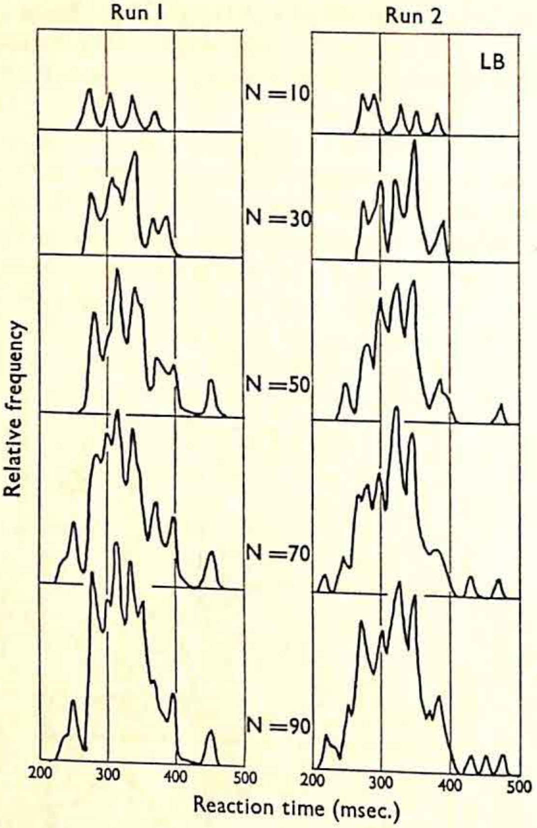


Examples of RT frequency distributions given by subjects RH, CW, NB, and LB in response to auditory clicks 8 dB. and 24 dB. above relative threshold. Each distribution is based on an N of 50. An upward deflection indicates increased frequency. A click was presented once every 4 sec. on the average, the interval varying from 3 to 5 sec.

Number of RTs grouped and periodicity. When the subjects responded for a long period of time and a large number of RTs were grouped in order to obtain the RT distributions, the periodicity tended to attenuate. To investigate the effects of recording time or the number of RTs grouped on the periodicity in the RT frequency distributions, distributions were recorded from three subjects after they had responded to 10, 30, 50, 70, and 90 light flashes. Two such sets of distributions from one subject are given in Figure 5.

It was found that the points in time when the greatest and least number of RTs occurred (the peaks and troughs of the distributions) remained stable for a relatively short period of time. The periodicity was most evident for the period of time required to obtain about 50 RTs under these conditions (about 7 min.); when more RTs were grouped, the periodicity became less evident. It may be noted that the

FIGURE 5



Effects of number of RT grouped on periodicity in RT frequency distributions obtained from subject LB. Subject responded to the onset of a 500 msec. light flash which was approximately 2 log units above relative threshold and subtended a visual angle of 17 min. A point 45° to the right of the flash source was binocularly fixated. The flashes were presented once every 8 sec.

latencies of the peaks and troughs of distributions obtained from successive replications (run 1 and 2, Fig. 5) varied across the period of time required to obtain the replications (15 min. under these conditions). This may also be noted in the data given in Figure 1.

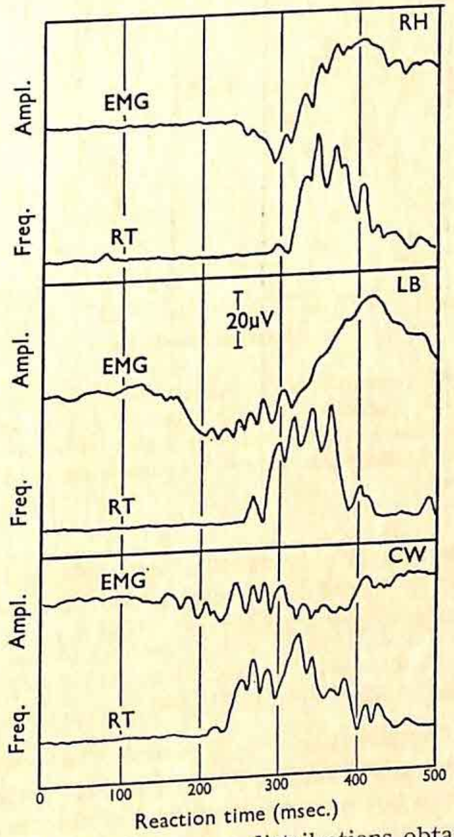
Electromyograms and periodicity

Hayes (1960) and Nightingale (1957) noted that electromyograms (EMGs), measured from the surface of the skin, had a modal frequency of approximately 40 cps. Since this is in accord with the frequency of the periodicity found in the RT distributions (see Fig. 2), the question was raised as to the relationship between these variables. Therefore, EMGs were measured in conjunction with the RT frequency distributions.

Electromyograms were measured by placing surface electrodes 4 cm. apart on the right forearm. The electrodes were bands 2 mm. wide and wrapped approximately 270 degrees around the perimeter of the forearm. Therefore, a measure of gross muscular activity was obtained. The skin was washed prior to attaching the electrodes to insure that skin resistance was less than 5,000 ohms. The muscle activity was amplified by an Offner type R Dynograph and then fed into the CAT for recording purposes. Band pass filters were used to attenuate activity above 22 cps. and below 5·3 cps. It should be noted that the use of the CAT in this system enabled muscular activity time-locked with the onset of the stimulus to be summed and thus be differentiated from the ongoing intrinsic muscular activity. The technique is the same as that used to sum evoked cortical potentials.

Summed EMG records and RT frequency distributions obtained by having subjects respond to 100 successive light flashes are given in Figure 6. In these and other records, the summed EMG records had a periodic nature very similar to that found in the RT distributions. In these data, the average period of the summed EMG activity was 20 msec. (50 cps.) and the RT frequency distributions was 23 msec. (43 cps.). It may also be noted that *(a)* the onset of the periodicity in the summed EMG record preceded the occurrence of the earliest reactions, and *(b)* there was no clear one-to-one correspondence between the peaks and troughs of the two sets of data.

FIGURE 6

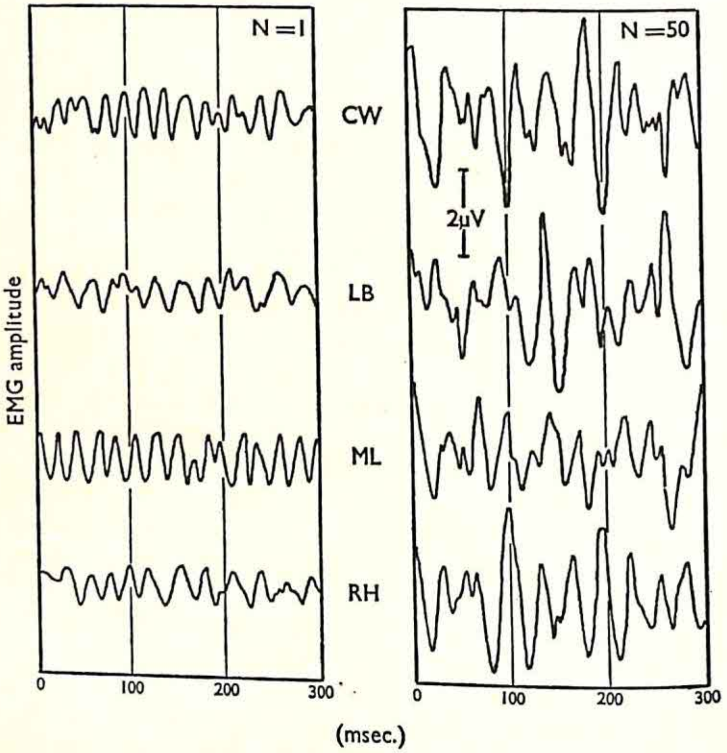


Summed EMG records and RT frequency distributions obtained from subjects RH, LB, and CW. Responses to 100 light flashes. The conditions are identical to those described in Figure 5.

The fact that some degree of periodicity was evident in the summed EMG records indicated that the EMG activity was partially time-locked with the onset of the stimulus or with the onset of the response which, in turn, was partially time-locked with the onset of the stimulus. This raised the question as to whether the periodicity in the summed EMG records was a function of the subject executing a response or if it was characteristic of summed intrinsic muscular activity.

Therefore, muscular activity was summed from subjects who were inducing a low level of muscular tension in their right forearm while fixating a bright flashing light but not reacting to it. Records resulting from single summations ($N = 1$) and 50 summations ($N = 50$) under these conditions are given in Figure 7. The single summations had a periodicity of 25–50 cps. which is consistent with the data reported by Hayes (1960) and Nightingale (1957). A periodicity of approximately the same frequency is evident in the summed EMG records. This indicates that some periodic muscular activity was synchronous with the presentation of the stimulus even when the subjects were not responding to the stimulus.

FIGURE 7



Single ($N = 1$) and summed ($N = 50$) EMG records obtained when subjects maintained a constant level of induced muscular tension. Subjects observed light flashes presented once per second but did not respond to them (see text).

In order to assess the modal period of the activity in the summed EMG records presented in Figure 7, the between-peak interval of the activity in six such records was measured for subjects CW and RH (3,000 msec. of activity and 300 summations for each subject). These measurements were made in the same manner as that used to assess the modal period of the RT frequency distributions. The resulting distributions of the total time occupied by each between-peak interval in the summed EMG data (see Fig. 2) indicated a modal between-peak interval of 25 msec. (40 cps.) for both subjects. This is the same modal between-peak interval found in the RT distributions. The rank-order correlation between the periodicity in the RT distributions and the periodicity in the summed EMG records was +0·46 and +0·53 for subjects CW and RH respectively ($p < 0{\cdot}025$).

It may be noted that the activity in the summed EMG records obtained when the subjects were not responding was not the result of the light flash *per se*. This conclusion is based on two observations. First, in control conditions where the light flash was withheld but the CAT was triggered in the regular manner, muscular activity summed in the same manner as when the light flash was presented. And second, the periodicity in the summed activity was evident immediately after the stimulus was presented; the expected latency between stimulus and response was absent. In light of these observations and the fact that the summed EMG activity had the same modal periodicity as raw EMG data, the summed EMG data appears to reflect the periodicity evident in the intrinsic activity of the muscle.

Discussion

The results of the present study indicate that the process or processes which underlay the periodicity in RT had the following characteristics:

(*a*) The process had a modal frequency of approximately 40 cps. or a period of 25 msec. as suggested by the data given in Figure 2.

(*b*) The process was, somehow, synchronous or time-locked with an external event (the stimulus or response) in order for its effects to be evident in the RT data. This may have come about in two ways. If the process was sufficiently stable, the rhythmically occurring stimulus may have occurred at a frequency which was a multiple of the periodicity of the process (the stimulus was time-locked to the process). Or, the process may have been "reset" or "triggered" by the onset of the stimulus or response (the process was time-locked to the event).

(*c*) Although the process was synchronous with stimulus onset, the latency between the onset of the stimulus and any given phase of the process varied over time. This may be noted in the distributions given in Figures 1 and 5, some of which were obtained under identical conditions. Latour (1961) similarly found that he had to compromise between the amount of information obtained, that is the number of RTs obtained in a given trial, and the stability of the subject.

(*d*) The process was involved in both visual and auditory RTs since the periodicity was evident in RT distributions obtained to visual (Figs. 1, 3, 5 and 6) and auditory (Fig. 4) stimuli.

The question is raised as to whether the periodicity in RT was the result of sensory, motor, or central processes. The relationship between the RT and EMG data noted above suggests that the periodicity was the result of motor activity. Past studies which have investigated a periodic component (20–50 msec.) in tremor and its relationship to movement lend further support to this suggestion (Fossler, 1931; Lansing, 1957; Tiffin and Westhafer, 1940; Travis, 1929; Travis and Hunter, 1931). Some evidence also may be offered that the periodicity was the result of sensory activity, at least for the visual RT data. Latour (1967) observed an ongoing oscillation of visual threshold with a period of approximately 27 msec. Since these fluctuations possibly resulted from physiological nystagmus ("eye tremor") which has a period ranging from 15–30 msec. (Latour, 1967; Ratliff and Riggs, 1950), they may have been the result of motor activity. Therefore, the present authors tend to favour the suggestion that the periodicity was the result of motor activity although the data are not conclusive in this respect.

It seems plausible that central mechanisms are involved in the periodicity in RT, regardless of whether it is the result of sensory or motor processes. Both eye movement and muscle tremor appear to be centrally controlled—eye movement by activity in the nuclei of the oculomotor, trochlear, and abducen cranial nerves (Gernandt, 1960), and muscle tremor by activity in the anterior horn cells of the spinal cord (Jung and Hassler, 1960). The fact that there is a periodic component in the electrical activity of the central nervous system with a period of 25–30 msec. (beta EEG activity) (Latour, 1967; Gaarder and Speck, 1967), and that the periodicity in RT is common to at least two sensory modalities (visual and auditory) lends tangential support to this conclusion. Further investigation of the relationship between RT, EMG, muscle tremor, EEG, and sensory threshold is necessary before a more complete attempt is made to assess the respective contributions of sensory, motor, and central processes to the periodicity in RT.

We thank Dr. R. G. Eason, L. Bonelli, and J. Hoke for their assistance in the execution of this study, and Dr. N. R. Bartlett for critically reviewing the manuscript. The first author was on a National Academy of Sciences-National Research Council Postdoctoral Associateship when this research was initiated. The manuscript is being submitted with the understanding that a right of reproduction for governmental purposes is reserved for the Naval Electronics Laboratory Center. The opinions and assertions contained herein are the private ones of the authors.

References

Augenstine, L. G. (1955). Evidences of periodicities in human task performance. In Quastler, H. (Ed.), *Information Theory in Psychology*. Glencoe: Free Press, pp. 208–26.

Eason, R. G., Oden, D., and White, C. T. (1967). Visually evoked cortical potentials and reaction time in relation to site of retinal stimulation. *Electroenceph clin. Neurophysiol.*, **22**, 313–24.

Fossler, H. R. (1931). Range and distributio [illegible] tremor frequencies. *J. gen. Psychol.*, **5**, 410–4.

Gaarder, K., and Speck, L. B. (1967). The quasi-harmonic relations of alpha and beta peaks in the power spectrum. *Brain Res.*, **4**, 110–2.

Gernandt, B. E. (1960). Vestibular mechanisms. In Magoun, H. W. (Ed.), *Handbook of Physiology: Vol. II*, Washington: American Physiol. Soc., p. 559.

Harter, M. R. (1967). Excitability cycles and cortical scanning: a review of two hypotheses of central intermittency in perception. *Psychol. Bull.*, **68**, 47–58.

Hayes, K. J. (1960). Wave analyses of tissue noise and muscle action potentials. *J. appl. Physiology*, **15**, 749–52.

Jung, R., and Hassler, R. (1960). The extrapyramidal motor system. In Magoun, H. W. (Ed.), *Handbook of Physiology: Vol. II*, Washington: American Physiol. Soc., p. 886.

Lansing, R. W. (1957). Relation of brain and tremor rhythms to visual reaction time. *Electroenceph clin. Neurophysiol.*, **9**, 497–504.

Latour, P. L. (1961). The eye and its timing. *Institute for Perception RVO-TNO, Rep. 1961–2* Soesterberg, Netherlands.

Latour, P. L. (1967). Evidence of internal clocks in the human operator. *Acta Psychologica*, **27**, 341–8.

Latour, P. L. and Bouman, M. A. (1961). A nonanalog time component in visual pursuit movements. In Rosenblith, W. A. (Ed.), *Sensory Communication*, New York: M.I.T. Press, pp. 798–803.

Nightingale, A. (1957). The analysis of muscle potentials by means of a Muirhead-Pametrada Wave Analyser. *Technique, J. Instrument Engineering*, **11**, 27–32.

Ratliff, F. and Riggs, L. A. (1950). Involuntary motions of the eye during monocular fixation. *J. exp. Psychol.*, **40**, 687–701.

Tiffin, J. and Westhafer, F. L. (1940). The relation between reaction time and temporal location of stimulus on tremor cycle. *J. exp. Psychol.*, **27**, 318–24.

Travis, L. E. (1929). The relation of voluntary movement to tremors. *J. exp. Psychol.*, **12**, 515–24.

Travis, L. E. and Hunter, T. A. (1931). Tremor frequency. *J. gen. Psychol.*, **5**, 255–60.

Venables, P. H. (1960). Periodicity in reaction time. *Brit. J. Psychol.*, **51**, 37–43.

Manuscript received 16th November, 1967.

EFFECTS OF NUMBER OF ALTERNATIVES ON THE PSYCHOLOGICAL REFRACTORY PERIOD

BY

LAWRENCE KARLIN and RICHARD KESTENBAUM

From Department of Psychology, New York University

The effects of changing the number of choices in the first response on "psychological refractoriness" were found to be partially consistent with intermittency theory but modifications may be needed to explain the results when the number of choices in the second response is increased.

INTRODUCTION

Current concepts of central intermittency in serial reaction time tasks derive from the work of Craik, Hick, Welford, Vince and Davis (see Bertelson, 1966 for an excellent review and analysis). Quantitative formulation was given by Welford (1952) who hypothesized that the central processes resulting from two separate successive stimuli could not overlap so that if the second stimulus arrived during the processing of the first it would be held in temporary storage until such processing was completed. In this way the theory purported to explain why, for short interstimulus intervals (I), the second response latency (RT_2) was increased above its "normal" value (RT_n). More specifically, RT_2 was predicted by Welford to vary according to the following equation:

$$RT_2 = RT_n + RT_1 - I \qquad (1)$$

This equation suggests that $RT_2 = RT_n$ when $I \geqslant RT_1$ but at these values of I there may still be a delay (Welford, 1959; Davis, 1956, 1957). So Hick (1948) and Welford (1952, 1959) hypothesized that when S_2 occurred while the first response (R_1) was being made, proprioceptive feedback from significant portions of R_1 such as its beginning and end may "capture" the central processing mechanism resulting in a further delay before S_2 could be processed. There are divergent views about the source of this secondary delay as well as other deviations from equation (1) that have been reported as I approaches and exceeds RT_1. Fortunately, the major purposes of the present experiment can be met by studying $RT_2 = f(I)$ where I is appreciably less than RT_1. In this region, the predictions of intermittency theory that $RT_2 = f(I)$ with unit negative slope should be minimally complicated by proprioceptive feedback and other factors such as variability in RT_1 as recently described by Bertelson (1967), and Welford (1967).

The present experiment aims to test some of the implications of Welford's intermittency theory as well as some alternative theories by studying $RT_2 = f(I)$ for different parametric values of RT_1. In the first part of this study the number of choices was held constant at two for R_2 while for each of three conditions the number of choices for R_1 was 1, 2, and 5, respectively. According to equation (1) or some modification of it such as described by Davis (1957), three parallel curves should be generated with differences in the corresponding ordinates of the three curves approximating the differences in RT_1. Each curve should break at the appropriate value of $I = RT_1$ to asymptote at $RT_2 = RT_n$. This break point may be modified if the proprioceptive feedback and other factors mentioned above are operative. In any case, the first prediction should be verified at least for those I short enough to separate the delay factor hypothesized in equation (1) from whatever complicating factors seem to be introduced when I approaches RT_1. The longer delay periods provided by the five- and two-choice conditions should be helpful in this regard.

Expectancy theory provides an alternative explanation of the so-called refractory phenomenon. This theory attributes the increase in RT_2 with short I to a lessened expectancy due to the lower probability of a second stimulus occurring just after the first because of the rectangular distribution of I. Since the same rectangular distribution of intervals is used in each condition, expectancy theory alone would not predict any differences in the ordinates of the three functions. The "strong" (see Bertelson, 1966) version of this theory ascribed *all* of the increase in RT_2 to temporal uncertainty when there is no event uncertainty. If the effects of temporal and event uncertainty are not equal or independent, the slopes of the three functions may change systematically. For example, the importance of event, relative to temporal uncertainty in producing delays may increase with the number of choices.

The predictions will also differ for Broadbent's perceptual sampling theory (Broadbent, 1958; Borger, 1963). This theory hypothesizes a temporal quantizing of the sensory input into successive samples of fixed duration. Perceptual processing of the input is thus not continuous but operates on these successive chunks. Broadbent asserts that increase in RT_2 is a function of when, during such a sample, R_2 occurs. He makes the additional assumption that R_1 closes the first sample and starts the new one. This theory, in its simplest form, would predict no change in the functions with different RT_1.

A second part of the present study will determine the effect of changing the number of alternatives in R_2 while holding those in R_1 constant. According to Equation (1), any change in RT_n produced by this manipulation should yield an equal change in RT_2 for all values of I.

Method

Apparatus and stimulus parameters. Series of pulses varying in frequency were recorded on magnetic tape. When played back through a read-out device these pulses triggered various relays at preprogrammed time intervals which then presented the stimuli and triggered three Hewlett–Packard chronoscopes measuring, respectively, I, RT_1, and RT_2. These three values were printed on paper tape at the end of each trial. Each trial started with a 40 msec. warning tone of 1,000 cps. followed by the first reaction stimulus (S_1) consisting of a 35 msec. flash of a digit by a Nixie tube seen through a polaroid filter. The second reaction stimulus consisted of either one of two 35 msec. tones of 600 or 3,000 cps. presented to the subject through a pair of headphones cushioned to minimize ambient noise. Two sets of micro-switches were actuated by pressing ¾ in. dia. buttons mounted in two panels, one for each hand. About 200 gm. pressure were required to actuate each key. The left hand set used in responding to the visual display consisted of 5, 2, or 1 buttons. The digits (1 through 5) displayed by the Nixie corresponded to responses by the fingers numbered in that order beginning with number 1 for the little finger. The right hand set used in responding to the tone always consisted of two buttons actuated by the index and middle fingers in response to the high and low tones, respectively. Except for the printer, the subject and the sound-proofed apparatus were in the same air-conditioned room. The fan of the air-conditioner was always kept running.

In each trial foreperiods between warning and first reaction stimuli were 1·7, 1·9, 2·1 or 2·3 sec. presented equally often in modified random order. Interstimulus intervals between first and second reaction stimuli measured between stimulus onsets were 90, 190, 290, 390, 490, 590, 690, 790, 890, 990, 1,050, or 1,150 msec. The intertrial interval was 5 sec.

Procedure. All subjects were run on five conditions, 5–2, 2–2, 1–2, 2–1, and 1–1, in which the first and second digits of each pair refer to the number of alternatives in the first and second reactions, respectively. For each condition, the subject attended one session a day usually on consecutive weekdays at about the same time each day. Each of the 12 interstimulus intervals was presented in modified random order 10 times for a total of 120 trials per session. The intervals were arranged so that they occurred about equally often throughout the series with runs of more than three eliminated. In any one session two different series of 60 trials each paired in all possible orders were used. After the 60th trial there was a 2-min. rest during which subjects removed the headphones and relaxed their hands.

Subjects were told what to expect and to respond as quickly as possible. They were not informed of errors or premature responses.

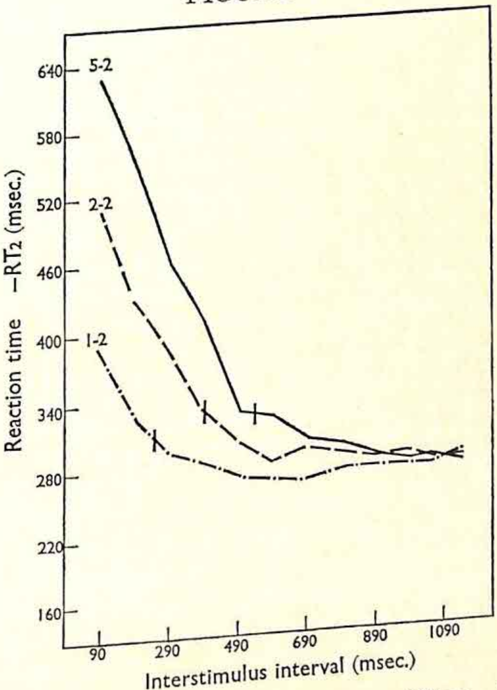
Practice. Prior to the present experiments subjects had been run for a total of 50 to 60 sessions in pilot experiments similar to the present one but with different visual stimuli, as well as in other refractory period experiments involving no choice in first and second reactions. An effort was made at the outset to counterbalance some of the conditions but this was not carried out because of subject attrition. The order of conditions for three subjects was: 5–2, 2–2, 1–2, 2–1, 1–1. Subjects attended seven or eight sessions for conditions 5–2 and 2–2 and 1–2, three to seven sessions for conditions 2–1 and two to four sessions on conditions 1–1.

Subjects. Six college students, paid one dollar per session, were selected from 10 on the basis of willingness to participate over a period of time and ability to pay attention during a session. Two subjects were dropped later, one because of frequent absences and the other because of excessively long and irregular RTs.

RESULTS AND DISCUSSION

Analysis of the results will be based in part on plots of RT, averaged over the four subjects, although the tables in each case will give individual data as well. Careful inspection of the individual results showed that in most instances overall means fairly represented each subject. Reference will be made to individual subjects as needed. The data for each subject are based on the mean of the last three sessions of conditions 5–2, 2–2, and 1–2. Trials were eliminated from the results in which either RT was less than 100 or more than 100 msec. longer than the next longest RT. For subject JB trials eliminated for these reasons averaged less than about 6 per cent. while for the

FIGURE 1



Mean RT_2 as a function of interstimulus interval by condition. Each point is the mean of approximately 360 observations derived from four subjects based on three sessions each.

other subjects it was less than 3 per cent. There was no discernible trend for the number of omitted trials to change with I.

Figure 1 compares RT_2 functions for the five-, two- and one-choice conditions (5–2, 2–2, 1–2) and is based on the results summarized in Table I. The figure shows that increase in RT_1 resulting from increasing the number of choices had the effect of markedly increasing the delay in RT_2 at the three or four shortest intervals.

TABLE I

MEAN RT IN MSEC. FOR BOTH RESPONSES (R_1 AND R_2) AS A FUNCTION OF INTERSTIMULUS INTERVAL BY SUBJECT AND CONDITION

		RT_1												
		Interval (msec.)												
Condition	*Subject*	*90*	*190*	*290*	*390*	*490*	*590*	*690*	*790*	*890*	*990*	*1,050*	*1,150*	RT_1
5–2	SS	434	458	448	500	462	470	448	458	488	476	472	471	465
	GR	447	476	456	464	459	467	467	459	464	460	465	471	463
	JB	594	584	602	598	538	592	588	606	592	545	578	571	582
	MS	521	507	499	532	543	530	529	533	531	531	535	520	526
	Obs. mean	499	506	501	524	501	515	508	514	519	503	512	508	509
2–2	SS	318	325	332	357	330	312	336	359	326	345	325	329	333
	GR	376	378	402	413	390	394	377	399	376	398	382	403	391
	JB	413	375	383	398	413	409	412	414	412	422	399	425	406
	MS	386	388	400	412	429	417	389	405	403	423	405	420	406
	Obs. mean	373	366	379	395	390	383	378	394	379	397	378	394	384
1–2	SS	214	233	221	216	228	229	221	221	221	219	221	231	223
	GR	203	212	212	209	211	211	212	218	215	217	208	215	212
	JB	257	250	242	259	258	241	250	247	244	238	247	246	248
	MS	264	286	292	291	278	282	290	295	297	296	261	271	284
	Obs. mean	234	245	242	244	244	241	243	245	244	242	234	241	242

		RT_2												
		Interval (msec.)												
Condition	*Subject*	*90*	*190*	*290*	*390*	*490*	*590*	*690*	*790*	*890*	*990*	*1,050*	*1,150*	RT_n
5–2	SS	573	512	433	378	311	325	289	286	252	263	260	267	263
	GR	609	526	425	368	318	291	276	279	270	271	273	268	271
	JB	701	610	544	467	350	365	351	339	338	314	330	300	315
	MS	637	531	440	422	354	323	296	287	280	266	258	276	267
	Obs. mean	630	545	461	409	333	326	303	298	285	279	280	278	—
	Theor. mean	688	595	490	413	290	279	—	—	—	—	—	—	279
2–2	SS	481	404	347	299	273	263	282	279	271	266	269	261	265
	GR	521	455	406	328	303	277	289	294	266	283	283	292	286
	JB	568	457	443	371	339	315	327	298	296	290	284	302	292
	MS	487	399	345	328	293	278	277	280	290	292	272	265	276
	Obs. mean	514	429	385	331	302	283	294	288	281	283	277	280	—
	Theor. mean	563	456	369	285	280	—	—	—	—	—	—	—	280
1–2	SS	352	307	267	264	252	252	246	244	260	252	244	258	251
	GR	412	354	299	304	270	270	267	272	258	261	264	276	276
	JB	391	342	300	277	282	275	283	296	288	295	291	306	297
	MS	397	322	319	303	288	281	269	272	287	285	297	297	293
	Obs. mean	388	331	296	287	273	269	266	271	273	273	274	284	—
	Theor. mean	421	334	279	—	—	—	—	—	—	—	—	—	279

Each individual entry is the mean of the last three sessions.

The implications of Figure 1 for Broadbent's sampling theory are negative since it would predict no change in RT_2 as a function of change in RT_1. Perhaps Broadbent did not intend to apply the theory so broadly, but until the relationship between the effects of amount of information being processed and sampling interval are known or hypothesized, the theory is not readily testable (see also Welford, 1959). Specifically, Broadbent does not predict what will happen if the processing time of the first stimulus is longer than the sampling interval. Possibly predictions from Broadbent's theory are limited to situations in which the sampling interval is greater than the central processing time. Of further interest and hinted at by Broadbent is to determine what factors influence the size of the sensory sampling interval. Evidence against Broadbent's sampling theory as described in the present study is also provided by the substantial correlations shown in Table II particularly in conditions 5–2 and 2–2. It could be argued, however, that at the shortest intervals some amount of correlation may be a result of the subject's state not having time to change.

TABLE II

Product Moment Correlations Between RT_1 and RT_2 by Interval, Condition and Subject

Condition	*Subject*	*Interval* (*msec.*)											
		90	*190*	*290*	*390*	*490*	*590*	*690*	*790*	*890*	*990*	*1,050*	*1,150*
5–2	SS	93	93	57	71	56	54	−21	27	31	16	53	−35
	GR	51	76	72	32	32	48	04	01	−01	−04	−12	02
	JB	92	81	95	83	83	60	−03	70	08	−12	−27	07
	MS	91	96	93	83	70	45	16	06	13	03	33	01
2–2	SS	85	68	77	90	31	29	23	03	−03	−07	13	−17
	GR	78	41	26	−23	20	21	22	13	−32	15	−04	06
	JB	92	68	66	57	83	64	17	−03	23	21	16	08
	MS	84	82	73	36	55	05	26	15	21	10	−32	−19
1–2	SS	−05	76	07	16	24	45	02	−25	35	−06	15	−01
	GR	44	−09	21	42	24	−02	−02	19	−06	15	10	15
	JB	70	53	33	71	45	−02	−14	24	14	44	37	50
	MS	55	51	64	39	28	24	03	16	17	18	−30	03

Expectancy theory, too (Bertelson, 1966), cannot readily explain the results of Figure 1 since the distribution of I is the same in each case. Temporal uncertainty may influence RT_2 within each condition but temporal uncertainty could not explain the consistent changes in ordinate values among conditions. In any case Adams's query "whether a necessary central delay time exists for the resolution of *event uncertainty*" (1962, p. 286) is answered affirmatively by the present data. Of interest, however, is the small but persistent tendency for mean RT_2 in conditions 1–2 to fall below the corresponding values in the other two conditions between 490 and 890 msec. Perhaps this result reflects the relatively greater role of tendencies described by Mowrer (1940) and Karlin (1959, 1966) to respond with shortest RT to the average rather than shortest or longest I. Davis (1956) discusses a similar tendency although the evidence for it is not clear in our condition 1–1 (see Fig. 3 below) which is more comparable to Davis's conditions.

In general the three curves of Figure 1 are approximately parallel for the three shortest intervals as predicted by intermittency theory. The condition 1–2 curve may differ from the other two curves in having a somewhat shallower slope although this kind of comparison may be somewhat vitiated because the differences between I begin to be too great at condition 1–2 to determine its slope with the same degree of precision that the other conditions afford.

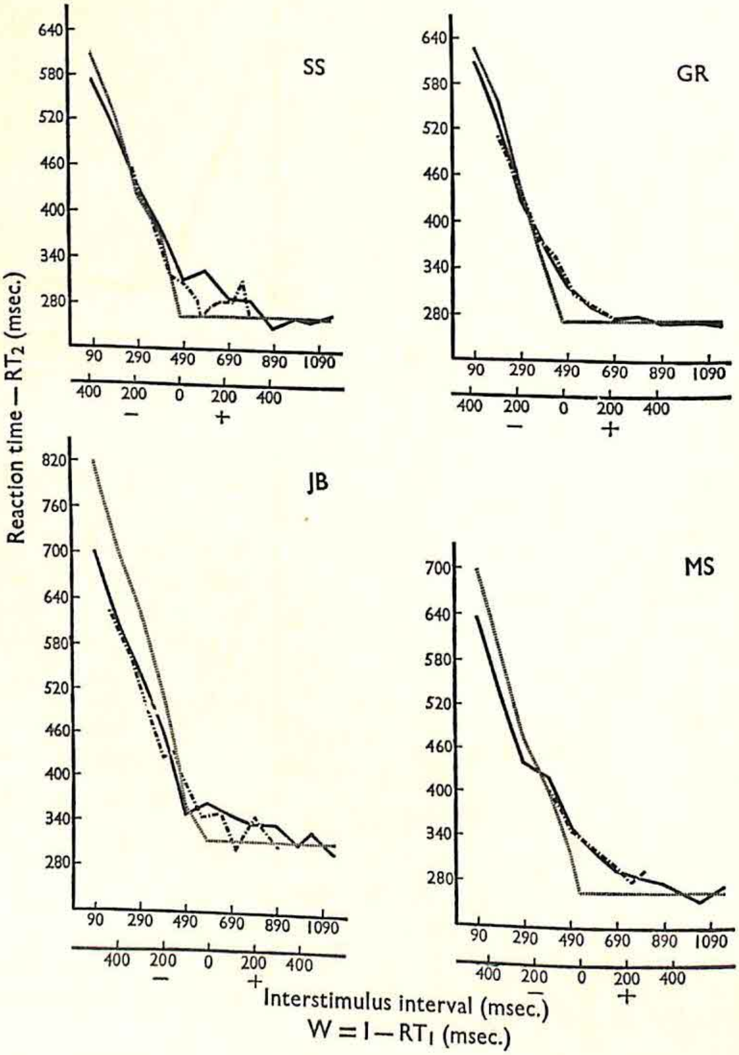
Intermittency theory would also predict that for a given I, increases in RT_2 and RT_1 should be equal providing I is short enough so that S_2 does not occur when R_1 is being made. Table III makes this comparison for the 90 and 190 msec. intervals by showing the increase in RT_2 expressed as a percentage of the increase in RT_1 for the

TABLE III

MEAN PERCENTAGE INCREASE IN RT_2 RELATIVE TO INCREASE IN RT_1 FOR ALL PAIRS OF CONDITIONS BY SUBJECT AND INTERSTIMULUS INTERVAL

	Interval (msec.)					
	90			190		
	Conditions compared			*Conditions compared*		
Subject	5–2 2–2	5.2 1–2	2–2 1–2	5–2 2–2	5–2 1–2	2–2 1–2
SS	79	100	124	89	91	105
GR	124	81	63	72	65	61
JB	73	92	113	73	80	92
MS	111	93	74	111	95	75
Mean	97	91	94	86	83	83

FIGURE 2A

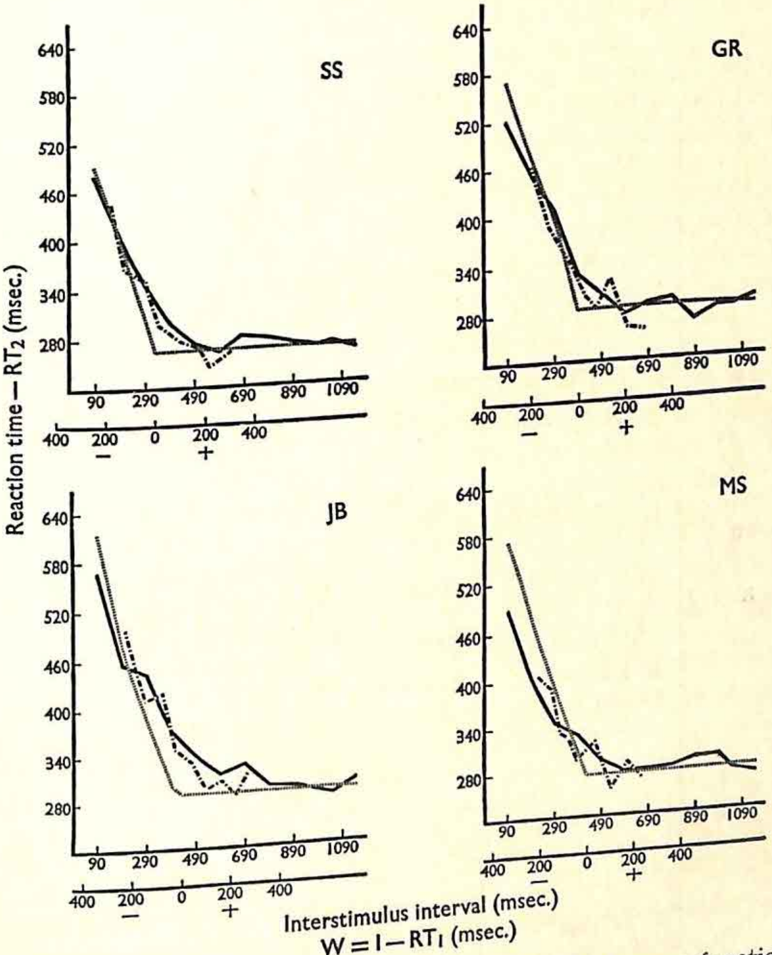


three possible pairs of conditions. It is clear that at both intervals the increase in RT_1 had a sizeable effect on RT_2. At the 90 msec. interval the means are fairly close to prediction from the kind of intermittency described by Welford and Davis. Partial confirmation of this theory is also suggested by the comparison at 190 msec. but the mean percentage changes in RT_2, while substantial, are appreciably lower.

None of the curves reaches asymptote (RT_n) at the predicted value of I shown for each case by the vertical bars in Figure 1 although condition 5–2 shows a definite change in slope near $I = RT_1$. The persistent delays may be due to feedback from the beginning of the response, averaging of subjects with different mean RT_1, or the effects of RT_1 variability (Bertelson, 1967; Welford, 1967).

To minimize these factors, Figure 2 shows the fit of equation (1) (broken line) to the individual curves in conditions 5–2 (Fig. 2A) and 2–2 (Fig. 2B) where RT_2 is

FIGURE 2B



Theoretical (broken line) and mean observed (solid line) RT_2 as a function of interstimulus interval and $w = I - RT_1$ (dotted line). A—condition 5–2; B—condition 2–2.

plotted both as a function of I (solid line) and the quantity, $w = I - RT_1$ (dotted line).

Considering $RT_2 = f(I)$ in condition 5–2 first, we note that the slope of the observed function is fitted rather well by theory for the first three intervals, although RT_2 within the range is consistently overestimated. This discrepancy might be due to using the entire RT_1 instead of subtracting out appropriate peripheral conduction and action times. On the other hand the discrepancy becomes positive and the slope tends to become shallower as I approaches and exceeds RT_1. The plots of condition 2–2 shown in Figure 2B may be described in much the same way although the slopes show a slight but consistent tendency to be a little less negative than those of condition 5–2 for short I.

Bertelson (1967) and Welford (1967) have argued that plotting $RT_2 = f(I)$ may not provide a valid test of intermittency theory when I is close to RT_1 because its variability will cause RT_2 to decrease more gradually, reaching RT_n only when I is appreciably greater than RT_1. Bertelson suggests measuring the independent variable to obtain an estimate of the actual overlap time between first and second reactions such as is provided by the quantity, $w = I - RT_1$. Condition 5–2, however, does not reveal much change in the slope of the function when w is used rather than I although condition 2–2 does show a small but consistent tendency for RT_2 to be reduced. Nevertheless, for all subjects $RT_2 > RT_n$ for the first one or two values of $w > 0$, therefore, in these conditions at least part and possibly all of the additional delay found when I is close to RT_1 cannot be ascribed to variability in RT_1. The possibility remains that this delay could be attributed to the effects of proprioceptive feedback described by Welford but the data are not conclusive.

Finally, the small but persistent differences found between 590 and 890 msec. in all three conditions which cannot be due to initial refractoriness or variability in RT_1 could also be due to proprioceptive feedback arising from the end of the response at different points in time for three conditions. These interpretations require further study with more precise measurement of when the beginning and end of the response actually occur (Bartlett, 1963).

In the conditions discussed above, R_2 was always a two-choice reaction. It was felt that event uncertainty when added to temporal uncertainty would provide an additional safeguard against effects introduced by any systematic strategy for the subject to group responses in some way (Halliday, Kerr and Elithorn, 1960). To check the effectiveness of this procedure and for other reasons (see Introduction), two other conditions were added, namely 1–1 and 2–1, in which there was only one choice in R_2. The effect of changing the number of choices in the second response on the relationship between mean RT_2 and I is shown in Figure 3. The results of the two new conditions (1–1 and 2–1) are summarized in Table IV. They support intermittency theory in that the difference between mean RT_2 is about equal to the corresponding differences between mean RT_1 at the 90 msec. interval. However, when the number of choices in the first reaction is held constant and that in the second is varied, the difference between RT_2 at $I = 90$ msec. is less than would be predicted by simple intermittency theory in each comparison. For example, Figure 3 shows that in conditions 2–2 and 2–1, the asymptotes (RT_n) differ by about 84 msec. Theoretically, this difference should obtain at all intervals. But at 90 msec. the difference is only 27 msec. In comparing conditions 1–2 and 1–1 a similar result is obtained, an asymptotic difference of 89 msec. being found along with a 39 msec. difference at the 90 msec. interval. If we correct the differences at the 90 msec. interval for their respective RT_1 (see Fig. 4) as intermittency theory would require, it leaves virtually no difference between the corresponding RT_2, making it even more difficult

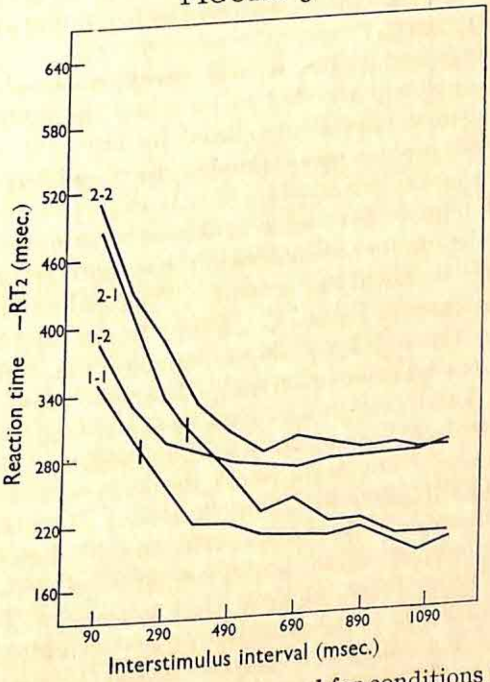
TABLE IV

MEAN RT FOR BOTH RESPONSES (R_1 AND R_2) BY INTERSTIMULUS INTERVAL FOR CONDITION 2–1 AND 1–1

		RT_1												
		Interval (msec.)												
Condition	*Subject*	*90*	*190*	*290*	*390*	*490*	*590*	*690*	*790*	*890*	*990*	*1,050*	*1,150*	RT_1
2–1	SS	322	318	313	340	333	312	322	320	310	317	305	319	319
	GR	366	394	363	368	383	380	395	377	370	356	375	397	377
	JB	385	387	376	394	426	382	367	364	370	348	372	394	380
	MS	357	348	380	347	354	352	354	345	350	368	332	338	352
	Obs. mean	357	362	358	362	374	356	359	351	350	347	346	362	357
1–1	SS	227	218	218	219	213	230	224	215	219	217	215	221	220
	GR	186	197	205	197	207	199	195	198	191	189	190	198	196
	JB	254	276	250	226	249	260	230	237	252	235	250	234	246
	MS	203	213	215	208	224	208	214	207	226	208	209	204	212
	Obs. mean	217	226	222	212	223	224	216	214	222	212	216	214	218

		RT_2												
		Interval (msec.)												
Condition	*Subject*	*90*	*190*	*290*	*390*	*490*	*590*	*690*	*790*	*890*	*990*	*1,050*	*1,150*	RT_n
2–1	SS	453	375	281	277	262	217	238	207	225	204	195	202	200
	GR	482	430	327	268	243	217	214	181	174	182	178	177	179
	JB	534	424	376	346	308	246	249	239	253	217	257	228	234
	MS	480	403	347	276	235	216	242	224	209	188	171	190	183
	Obs. mean	487	408	333	292	262	224	236	213	215	198	200	199	—
	Theor. mean	466	371	267	199	—	—	—	—	—	—	—	—	199
1–1	SS	326	258	233	209	201	196	189	192	189	202	183	188	191
	GR	359	320	296	226	202	205	192	187	206	188	186	197	190
	JB	388	340	271	231	260	228	245	238	259	219	213	241	224
	MS	322	281	224	205	203	188	179	179	169	157	158	152	156
	Obs. mean	349	300	256	218	216	204	201	199	206	191	185	194	—
	Theor. mean	317	226	190	—	—	—	—	—	—	—	—	—	190

FIGURE 3

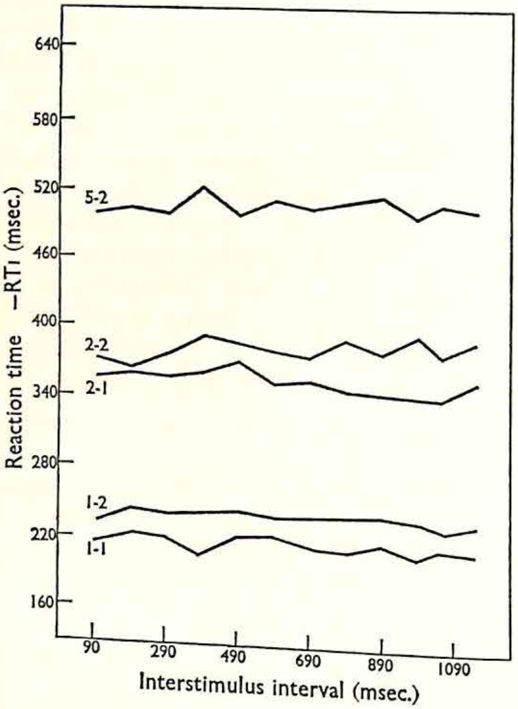


Mean RT_2 as a function of interstimulus interval for conditions 1–1, 1–2, 2–1, and 2–2.

to explain why the asymptotic differences between one-and two-choice RT_2 in either condition are not maintained at 90 msec. These findings apply to each subject without exception and are not predictable from equation (1) since it assumes (for short I) that the delay in RT_2 is only a function of the central processing time of the first reaction and that once this is completed the remaining latency should be "normal."

Possibly these results could be attributed to some kind of grouping strategy in which the subject delays his first response to determine what the second response will be. Such a strategy might be revealed by changes in RT_1 with I (Borger, 1963). Figure 4 plots the relevant results for all five conditions and reveals no significant finding in this regard, all functions being essentially flat. However, for conditions 1–1, 1–2, 2–1 and 2–2 there is a consistent tendency for increase in the number of choices in R_2 to lengthen RT_1 when the number of choices in R_1 is held constant both at one and at two. This result may mean that the subject tends to wait a fixed amount of time before he responds to the first stimulus, possibly because of some general set factor related to task complexity (Bertelson, 1967) or perhaps to find out which choice of second stimulus will be presented. The waiting time, however, does not seem to be large enough for the latter purpose, averaging under 30 msec. over all intervals in each comparison when it would need to be more than 90 msec. if the subject geared his waiting to the shortest inter-stimulus interval. Conceivably the shorter RT_1 for conditions 1–1 and 2–1 are themselves delayed so that 30 msec. may be an *additional* waiting time. If these differences in RT_1 were related to the results of Figure 3, quantitative relationships might be revealed by individual differences. A

FIGURE 4



RT_1 as a function of interstimulus interval for each condition good comparison of this kind is provided by SS and MS for conditions 1–1 and 1–2. From Tables I and IV the mean RT_1 differences for these two subjects are 3 and 72 msec., respectively. Despite this large difference (consistent over all I), both subjects clearly yield the relationships between the two functions shown in Figure 3 where we might expect that SS would hardly show these relationships.

Another possibility to account for the results shown in Figure 3 is that the centra channel while occupied with the first response may be able to process some kinds o information from the second response. Thus the two pair of functions may converge because in each pair conditions 1–2 and 2–2 produce only conditionally refractory delays at short I.

Perhaps the basic processes involved when RT_2 is simple cannot overlap with the refractory period produced by the first stimulus but some part of the additional selective processes involved in disjunctive reactions can. If we attribute the full delay in conditions 1–1 and 2–1 to refractoriness then the shortening of the delays in the other conditions using disjunctive RT_2 might be a result of such overlap and it is for this reason that the two pair of curves shown in Figure 3 approach each other at $I = 90$ msec. Specifically, RT_2 may not increase as much in conditions 2–2 and 1–2 because refractoriness is not as complete as in the case of the other two conditions.

A partial test of this rationale would be given by increasing the number of choices for RT_2 beyond that used in the present study. With increasing number of choices and the corresponding increase in RT_n, the refractory effects introduced by the first reaction might be increasingly overestimated by equation (1).

If processing of both reactions does overlap to some extent then RT_1 might also be lengthened with increase in the number of choices in RT_2 because of the additional load. Such a possibility offers an alternative explanation to the one given above (Fig. 4) for the small increase in RT_1 which obtains for both pairs of conditions in three of the four subjects.

This investigation was supported by Public Health Service Research Grant MH 07253 from the National Institute of Mental Health, Public Health Service.

The writers express their appreciation to Mr. M. Hamburg and Miss Patricia Odintz who assisted in collection and analysis of the data.

REFERENCES

ADAMS, J. A. (1962). Test of the hypothesis of psychological refractory period. *J. exp. Psychol.*, **64**, 280–7.

BARTLETT, N. R. (1963). A comparison of manual reaction times as measured by three sensitive indices. *Psychol. Rev.*, **13**, 51–56.

BERTELSON, P. (1966). Central intermittency twenty years later. *Quart. J. exp. Psychol.*, **18**, 153–63.

BERTELSON, P. (1967). The refractory period of choice reactions with regular and irregular interstimuli intervals. *Acta Psychol.*, **27**, 45–56.

BORGER, R. (1963). The refractory period and serial choice reactions. *Quart. J. exp. Psychol.*, **15**, 1–12.

BROADBENT, D. E. (1958). *Perception and Communication*. London: Pergamon Press.

DAVIS, R. (1956). The limits of the "psychological refractory period." *Quart. J. exp. Psychol.*, **8**, 24–38.

DAVIS, R. (1957). The human operator as a single channel information system. *Quart. J. exp. Psychol.*, **9**, 119–29.

DAVIS, R. (1959). The role of "attention" in the psychological refractory period. *Quart. J. exp. Psychol.*, **11**, 211–20.

HALLIDAY, A. M., KERR, M., and ELITHORN, A. (1960). Grouping of stimuli and apparent exceptions to the psychological refractory period. *Quart. J. exp. Psychol.*, **12**, 72–89.

HICK, W. E. (1948). Discontinuous functioning of the human operator in pursuit tasks. *Quart. J. exp. Psychol.*, **1**, 36–51.

KARLIN, L. (1959). Reaction time as a function of foreperiod duration and variability. *J. exp. Psychol.*, **58**, 185–91.

KARLIN, L. (1966). Development of readiness to respond during short foreperiods. *J. exp. Psychol.*, **72**, 505–9.

MOWRER, O. H. (1940). Preparatory set (expectancy): Some methods of measurement. *Psychol. Monogr.*, **52**, (2, Whole No. 233).

WELFORD, A. T. (1952). The psychological refractory period and the timing of high-speed performance—a review and a theory. *Brit. J. Psychol.*, **43**, 2–19.

WELFORD, A. T. (1959). Evidence of a single channel decision mechanisms limiting performance in a serial reaction task. *Quart. J. exp. Psychol.*, **11**, 193–210.

WELFORD, A. T. (1967). Single-channel operation in the brain. *Acta Psychol.*, **27**, 5–22.

Manuscript received 17th July, 1967.

THREE KINDS OF ERROR-SIGNALLING RESPONSES IN A SERIAL CHOICE TASK

BY

PATRICK M. A. RABBITT*

From the Medical Research Council, Applied Psychology Research Unit, Cambridge

During serial self-paced choice response tasks mean reaction times (RTs) for responses which are made in order to correct errors are faster than mean RTs for other correct responses. Experiment I showed that subjects can accurately correct errors in a four-choice task by making the response which they should have made, even though they are given no indication that an error has occurred. Experiment 2 showed that subjects correct their errors faster and more accurately when they use correction procedure than when they make a common response to all errors. The implication that subjects can correct errors because they know what response they should have made allows some comments on the constraints which must be met by various models which have been proposed to explain error-correction.

INTRODUCTION

In continuous-performance self-paced choice-response tasks the experimental equipment frequently gives the subject no indication whether any response which he makes is correct or wrong. Nevertheless subjects can very efficiently detect any errors which they may make (Rabbitt, 1966*b*, 1967). This implies that a subject can make a response, which he presumably believes to be correct at the moment when he initiates it, but which he shortly realizes to be wrong without any additional information from the experimenter or the equipment on which he is tested. Obviously the subject must make some check or comparison in order to detect a discrepancy between what he has done and what he ought to have done. Any model undertaking to describe his performance must specify what is checked or compared against what. If this can be stated we shall also have information relevant to an associated problem: Whether the subject's detection of an error implies that he knows what he ought to have done, or whether it simply means that he is aware of some discrepancy between his act and his intention, but cannot define the exact nature of the mismatch.

The necessary question is, therefore, whether a subject can correct his error by specifying the response which he should have made when correction procedure involves a choice between several responses. Some previous experiments examining error-correction by correction-procedure have used two-choice tasks, where the correcting response was always defined for the subject as the response which he did not make (Rabbitt, 1966*b*). Other experiments involving more than two response-choices have unfortunately been made with experimental equipment which always repeated any signal to which the subject had wrongly responded (Burns, 1965; Rabbitt, 1966*a*). Experiment I was therefore made to avoid both these limitations; that is, to discover whether subjects carrying out a serial self-paced four-choice task could correct their errors by specifying which of the three possible responses remaining after they had committed an error was the one which they should have made.

EXPERIMENT I

Subjects. Twelve Royal Navy ratings aged from 18 to 25 years (mean age 20·4) served as subjects.

---

* Now at Institute of Experimental Psychology, University of Oxford.

Apparatus

A stimulus presentation and reaction-timing apparatus (SPARTA) was programmed to present the digits 1 through 8 in random order, one at a time, on a "digitron" GRIOJ numerical display tube. These signals were answered by pressing one of four response keys inset into a desk so that the subject could conveniently depress them while resting his palms on a flat surface. The appearance of either one or two was answered with the left middle finger; three and four were answered with the left forefinger; five and six were answered with the right forefinger and seven and eight were answered with the right middle finger.

A sequence of responses began with the presentation of one of these digits, which the subject was required to immediately answer by pressing the appropriate key. Any response, correct or wrong, had precisely the same effect: that is, the display changed within 20 msec. of a switch closure, and the elapsed time since the subject's last response was read off to within 0·01 sec. A code for each response was punched out on five-channel paper tape. This specified the signal presented, the key which the subject pressed in answer to it, and the elapsed time since the last response.

The subject was instructed to run through sequences of 301 signals and responses, answering each new signal in turn as fast and accurately as possible.

Procedure

Ten different random sequences of 301 signals were programmed. Within each sequence all transitions between signals including repeats, occurred with as near equal frequency as possible. Subjects were run through two new sequences on each of five successive weekday mornings. Instructions before each of these sequences urged subjects to minimize their errors, and to correct any errors which did occur as quickly as possible by immediately making the response which should have been made.

Since the display changed within 20 msec. of *any* response, correct or wrong, an error also automatically triggered a new signal. The subject had to ignore this signal in order to make an appropriate error-correcting response (ECR). By chance, on a quarter of occasions, the ECR would be appropriate to the new signal triggered by the error. In order to be sure that such responses were intended as ECRs subjects were instructed to pause from 5–10 sec. after each error-correction in order to confirm their recognition of the error by recording on SPARTA output a RT which could never otherwise occur during the task.

Results

Output tapes for each subject's last two runs of 301 responses were teleprinted and analysed by eye.

Incidence of errors. One subject made no errors, and his results were therefore discarded. The remaining 11 subjects averaged 3·2 per cent. of errors. Using the occurrence of a 5–10 sec. pause as a criterion subjects acknowledged 74·1 per cent. of the errors which they made. One subject on one occasion, acknowledged an error which he did not commit (i.e. a false positive).

Error-correcting responses occurred in two different contexts: any given ECR might or might not be appropriate to the new signal triggered by the error. Summing over all subjects the incidence of these two types of ECR was very much what would have been expected by chance (ECRs were appropriate to post-error signals on 20·2 per cent. of occasions). Using the occurrence of a 5–10 sec. pause as a criterion there was no significant difference in the speed or reliability of error-detection whether the ECR was appropriate to the post-error signal or not (Mann-Whitney U, $p > 0{\cdot}1$, two-tailed).

When subjects detected errors they very reliably made the response which they should have made. Two subjects, each on one occasion, followed an error by making a response which was neither an appropriate ECR nor a response appropriate to the post-error signal. In both these cases the two successive errors were followed by a 5–10 sec. pause signalling recognition that an error had occurred. Otherwise all correction responses were accurate. Mean RTs and ECRTs are set out for comparison in Table I below.

TABLE I

Mean Correct RT and Error Correcting RT in a Four-Choice Serial Self-Paced Task with Error-Correction Procedure

	Mean (msec.)	σ
Overall Correct RT	532	114
Error Correcting RT	404	184

Mean ECRTs and Correct RTs for each of the 11 subjects were compared by Wilcoxon tests. ECRTs were significantly faster ($p < 0{\cdot}01$, two-tailed).

Discussion

As in previous experiments, although subjects were given no feedback from the experimental equipment, they almost always knew when they had committed an error (Rabbitt, 1966*b*, 1967). The results show that subjects did not merely know that some, unspecified discrepancy had occurred—since they could reliably specify which of three remaining responses they should have chosen. Evidently they had *complete* information as to which was the correct response, rather than *partial* information as previously suggested by Burns (1965) and Rabbitt (1966*a*).

The fact that subjects could correct their errors in Experiment 1 does not mean that they need always monitor their performance in this way. It remains possible that the amount of information which subjects have to process in order to specify the response which they ought to have made is greater than the amount which must be processed merely to discover that some (unspecified) failure in performance has occurred. If the process of error-*detection* is less complex than the process of error-*correction*, it is possible that subjects who are required to make a single common response to all errors should do so faster than subjects who have to correct errors by specifying the response which they should have made.

A previous study (Rabbitt, 1967) required subjects performing a version of Experiment 1 to make the same common detection response (CDR) to all errors by simultaneously pressing two keys on a console with their two thumbs. Since this was an anatomically clumsy response it is not interesting that their mean error detection times (484 msec., σ149) were slower than the error-correcting responses (ECRs) recorded in Experiment 1. It remained necessary to discover whether subjects could signal recognition of their errors faster when they used a CDR or when they used ECRs, under conditions when these were anatomically comparable responses. Experiment 2 was made to do this.

Apart from the question of their anatomical convenience, CDRs may be of two different kinds: (1) the CDR may be a response used for no other purpose than to signal the recognition of an error (task-external CDR) or (2) the CDR may be a response which is also made to some signals presented during the task (task-response CDR). The second class of (task response) CDRs allows an extra comparison because they can, by chance, fall into two contexts. A given task-response CDR, made to correct a given error, may also be the response which the subject ought to have made (i.e. an ECR-equivalent), or it may not. The use of task-response CDRs thus offers an opportunity of comparing anatomically identical ECR equivalents and other CDRs made under the same experimental instructions. Since the experiment demands that the incidence of errors in choice-response tasks should be kept as low as possible, the most convenient situation is a two-response task where we may expect to accumulate sufficient responses in each category for a valid comparison.

Experiment 2 therefore undertook to compare three kinds of responses to errors: ECRs, task-external CDRs which were anatomically similar to the ECRs, and task response CDRs—which might or might not be ECR-equivalents.

EXPERIMENT 2

Apparatus

SPARTA was again used, with the 10 programmes of signals made up for Experiment 1. All subjects carried out the same basic task: they responded to all digits one through four by pressing a response key set under their left forefingers and responded to digits five through eight by pressing a key under their right forefingers.

Subjects and procedure

Thirty-six Royal Navy ratings aged from 18·1 to 29·4 years (Mean 23·0) were individually run through two new sequences of 301 signals and responses on each of five successive weekday mornings. General instructions to respond as fast and accurately as possible were identical for all subjects. Three groups of 12 subjects were urged to detect and signal all their errors, as fast as possible, in three different ways.

ECR group. All 12 subject signalled their recognition of an error by immediately making the response which they should have made.

Task external CDR group. Six subjects signalled recognition of all errors by pressing a key beneath their left middle finger and six subjects signalled errors by pressing a key beneath the right middle finger. Neither of these responses was ever used during the task.

Task-response CDR group. These subjects signalled error-recognition by making, to all errors, a response which was also appropriate to half the signals presented on the display. Specifically, six subjects signalled all their errors by responding with the left forefinger on the key used to answer digits one through four, and six subjects responded with the right forefinger on the key used to answer digits five through eight.

Before every run all subjects were urged to signal their error-recognitions as quickly and accurately as possible, and were reminded of the error-signalling procedure required of them. Interim checks on the data allowed subjects to be told when they were missing errors or making wrong error-signalling responses.

Following every error-signalling response subjects were required to pause for 5–10 sec. before resuming the run. This was done so that error-recognition could be confirmed in situations where SPARTA output left the subject's intentions ambiguous.

Results

SPARTA output tapes for each subject's ninth and tenth runs of 301 signals and responses were teleprinted and analysed by eye.

Incidence of errors

Percentages of errors made and detected by each group are given in Table II. Analysis of variance showed no significant differences between groups.

Valid and invalid error-signalling responses. Two kinds of invalid error-signalling responses were possible. Subjects might make an error-signalling response although no error had occurred (False positive) or they might make a different response from the one which they had been instructed to make. Table II shows that false positives were very rare for all groups, and that only groups making CDRs made other, invalid error-signalling responses. These invalid responses were all of the same kind: subjects did not make the CDR but, instead, made the response which they should have made, (i.e. the ECR). Such contingencies could be identified by the occurrence of a pause indicating the subject's belief that he had made a correct error-recognition response.

These wrong correction responses were significantly (Wilcoxon, $p < 0{\cdot}05$) more common among the task-response CDR group (9·1 per cent. of CDRs) than among the

TABLE II

INCIDENCE OF ERRORS AND ERROR-DETECTIONS IN A TWO-CHOICE SERIAL SELF-PACED TASK

	Errors as percentage of total responses	*Percentage of total errors detected*	*Percentage wrong detection responses*	*"Detection" in absence of an error*
Error Correction Procedure (Group 1)	4·78	71·4	0	2
Task-response CDR (Group 2)	5·56	71·2	9·1	4
Task-external CDR (Group 3)	5·82	78·1	2·0	1

Three separate groups of 12 subjects used error correcting procedure and two types of common error detection responses.

task-external CDR group (2·0 per cent. of CDRs). In the task-response CDR group no subject ever made the wrong error-signalling response when the CDR required of him happened to be an ECR equivalent. In the task-response CDR group ECR equivalent and non-equivalent CDRs were equiprobable, and 10 out of 12 subjects made the wrong CDR. This effect must therefore be judged significant (Sign test, $p < 0·01$). Error signalling responses made by groups 1 and 2 might or might not be appropriate to the new signal triggered by the preceding error. The incidence of these events was approximately as would be expected by chance, and did not differ significantly between ECR and task-response CDR groups. Wilcoxon tests within each of these groups showed that (i) The incidence of valid error-recognitions, judged by the presence of a 5–10 sec. pause was not significantly different for each of these classes of events. (ii) RTs of error-signalling responses were the same for both classes of events, within each group.

RTs of error-signalling responses

Mean correct RTs and error-signalling RTs are set out in Table III.

TABLE III

MEAN CORRECT RT AND ERROR DETECTION RT

Type of error detection/correction response	*Mean correct RT* (msec.)	*Error detection/correction RT* (msec.)
Correction Procedure (ECR) (12 subjects)	435 ($\sigma = 93$)	396 ($\sigma = 152$)
Task response CDR (12 subjects)	428 ($\sigma = 104$)	
(1) Overall		676 ($\sigma = 248$)
(2) ECR equivalent		584 ($\sigma = 189$)
(3) Not ECR equivalent		758 ($\sigma = 265$)
Task external CDR (12 subjects) ..	437 ($\sigma = 84$)	467 ($\sigma = 170$)

The three groups of 12 subjects carrying out a two-choice serial self-paced task with different error-signalling procedures: (*a*) Error Correction Procedure (*b*) a task-response Common Detection Response and (*c*) a task-external Common Detection Response).

Analysis of variance showed no significant difference between mean correct RTs in the three groups. A separate analysis of variance showed that RTs of error-signalling responses were significantly different between groups ($p < 0·01$). S^2

showed that ECRTs were significantly faster than task-external CDRTs ($p < 0{\cdot}05$), which were in turn significantly faster than task-response CDRTs ($p < 0{\cdot}01$). Wilcoxon tests were made to compare Error-signalling RT against mean Correct RT within each group. In the ECR group Error-signalling RTs were significantly faster than mean Correct RTs ($p < 0{\cdot}01$, two-tailed). In contrast both task-control and task-response CDRTs were significantly slower than the mean Correct RTs ($p < 0{\cdot}05$ and $p < 0{\cdot}01$, two-tailed).

Task-response CDRs might or might not be ECR equivalents. As we have seen, subjects in the task-response CDR group tended to make ECRs, even when this contradicted instructions given over four previous days of practice. Ignoring such incorrect error-signalling responses, mean latencies were computed for correct ECR equivalent CDRs (Mean $= 584$, $\sigma = 189$) and for other CDRs (Mean 758, $\sigma = 265$). ECR equivalents were significantly faster (Wilcoxon, $p < 0{\cdot}01$, two-tailed). A series of Wilcoxon tests compared these against RTs for other classes of responses with the following results:

(i) Both types of task-responses CDRs were slower than mean correct responses ($p < 0{\cdot}01$, two-tailed).

(ii) Both types of task-responses CDR were slower than ECRs obtained from the ECR group ($p < 0{\cdot}01$, two-tailed).

(iii) Both types of task-response CDR were slower than task-external CDRs ($p < 0{\cdot}01$, two-tailed).

Discussion

Relevant features of the results of both experiments may be briefly summarized:

(1) In Experiment 1 subjects reliably recognized their errors and could nearly always specify which of three other responses they should have made.

(2) In Experiment 2 ECRs were faster than either task-external or task-response CDRs. "Real" ECRs are significantly faster than task response ECR equivalents.

(3) In Experiment 2 task-external CDRs were faster than task-response CDRs.

(4) Task-response CDRs were faster when they were ECR equivalents than when they were not.

(5) Consistent with (4) is the finding that subjects made task-response CDRs relatively inaccurately because they frequently made ECRs instead. The reverse confusion never occurred.

All these points support the generalization that subjects signalled their errors most quickly and accurately when they were allowed to make the response which they should have made. Points (3) and (4) raise another issue: task-response CDRs which are *not* ECR equivalents are relatively slow and inaccurate. Two reasons for this seem possible:

(*a*) There may be a strong competing tendency to make an ECR.

(*b*) In a two-choice task, a response which is *not* an ECR is, by definition, a repeat of the error. It may be that subjects are inhibited from repeating a response which they have just discovered to be wrong.

A direct experimental choice between (*a*) and (*b*) would require data from choice-response tasks with three or more possible responses, in which subjects make task-response CDRs. Here three classes of CDR could be compared: ECR equivalents, error-repetitions and other CDRs. Meanwhile an available comparison strongly supports hypothesis (*b*). Task-external CDRs were significantly faster than the task-response CDRs ($p < 0{\cdot}01$, Mann-Whitney, two-tailed). Thus, if non-ECR equivalent CDRs were slow *only* because of a competing tendency to make an ECR (hypothesis (*a*)) this should be equally true of task-external CDRs. Since task-external CDRs are faster it is most likely that task-response CDRs are slowed because they require the subject to repeat a response which he has just discovered to be wrong.

The central point to emerge from both experiments is that subjects not only know when they have made an error but apparently also know which response they should have made (Experiment 1). Error-detection, where the correct response need not be specified, does not take any longer than error-correction where the subject must not only signal that he has made an error, but also indicate what he should have done (Experiment 2). A previous discussion of the process of error-detection (Rabbitt, 1967) considered three possible models for the process by which a subject might detect any errors he makes without reference to any external source of information. We may now consider to what extent each of these models implies that a subject can correct an error because he knows what the appropriate response should have been.

Let us call the entire (unknown) sequence of C.N.S. events intervening between the onset of a signal and the completion of an overt response to it the "S → R computation." There are, then, at least three ways in which the accuracy of a given S → R computation may be checked.

(A) The subject may recompute the entire S → R computation one or more times. This would imply that when he detects an error the datum against which he checks his response is a solution specifying some other response which he did not make. The subject thus knows that he has made an error because he can specify another response. Note that if the subject is assumed to know *which* of two or more rival S → R computation solutions was correct (and not simply that they gave different results) we must introduce additional assumptions.

(B) The subject may only recompute part of an S → R computation. It would follow that only errors in the portion recomputed could be detected. If only the initial stages of the S → R computation are duplicated, the fact that the subject recognizes a discrepancy between separate computations need not imply that he knows which response he should have made. On the other hand, if it turns out that he *does* know what response he should have made, we must necessarily assume that the terminal stages of the S → R computation have been duplicated. In this case (as in A above) we will also have to explain how the subject apparently knows which of the conflicting solutions was correct.

Both models A and B assume that a solution to one S → R computation (which has been emitted as a wrong response) is checked against the result of some other, equivalent computation. A mismatch between two independent solutions would tell the subject that one of his computations was wrong. But we have seen that the subject knows more than this—he knows which is the correct solution. Clearly additional assumptions must be made for either A- or B-type models to be tenable. There seem to be three possible assumptions between which to choose.

(i) We may suppose that a subject who detects a mismatch between the solutions of separate S → R computations always prefers the solution which specifies a different response to the one which he had made. This would not be a good strategy on present evidence. Since 2–5 per cent. of responses are wrong, we must infer that this is true of S → R computations in general, and so that it must also be true of recomputed solutions. Thus we might expect that on 2–5 per cent. of cases the subject would think that he was wrong when he was not. In other words he must make as many false error-detections as he makes errors. In fact the probability of a false error-detection was about 0·0001 in these experiments. We must accordingly either reject the idea that the subject detects his errors by recomputation, or find some other modification of hypotheses A and B which explains why recomputed solutions are far more reliable than the responses which are checked against them.

(ii) Any given S → R computation may be supposed to be duplicated in parallel, not once but many times (Rabbitt, 1967). All these separate computations may be subject to random noise. Because of the random distribution of noise between (as well as within) computations some of these computations are likely to terminate in solutions before others. For the same reason, some of these computations may terminate in false solutions while others do not. A subject may achieve relatively reliable computation in noise by waiting for a consensus of solutions from several independent computations. He may act upon a small sample of these solutions at some risk of making an erorr, but possibly with an advantage in speed. Whether he responds prematurely or not, all these processes will eventually terminate providing a final, relatively reliable "consensus" which may serve as a check for the response he has chosen.

(iii) A similar model might assume that a subject can continue processing perceptual evidence after the moment in time at which he makes an overt response. Should the "premature" response be in error the subject will shortly realize this as his accumulation of evidence progresses towards a more reliable conclusion. This model has underlying assumptions common to recent models for choice-reactions (e.g. Laming, 1968), which assume that errors are, in some sense, "premature" responses based on inadequate sampling of perceptual evidence.

An interesting variant of such a model might predict that the subject continues to input and process information from the display after the moment in time when he has initiated his response. In the present experimental situation it may be argued, for instance, that because the display was only switched off 20 msec. after the subject has actually closed the response key he might, during this time, re-identify the signal and then initiate a second S → R computation. Since error-correction RT is shorter than mean correct RT this is hardly likely unless we adopt a curious but necessary hypothesis that a repeated S → R computation is undertaken faster on its second run. There is no reason, of course, why the subject should not re-sample the display at a moment in time before his overt, wrong response has been emitted. An empirical test of this hypothesis requires a study to determine whether subjects can accurately employ correction-procedure when they must respond to signals which are presented for such brief durations that the possibility of "double sampling" of the display is eliminated.

It seems that neither model A nor model B is satisfactory without some additional assumptions to explain the reliability of the recomputed solutions against which emitted responses are checked. Among the possible assumptions considered we can eliminate (i) above, but the data do not allow us to choose between (ii) and (iii). A third model for error-detection offers a more simple explanation of the data.

(C) Logically distinct from the idea of recomputation is a suggestion made by Miller, Galanter and Pribram (1962). Some part of the S → R computation may leave the subject in possession of a C.N.S. "program" which specifies the response he should make. Feedback from the response which he actually makes may then be compared against this "program." Such a comparison between an "act" and an "intention" can only detect errors in the execution of the act. A failure to form a correct intention (such as might result from perceptual misidentification of a signal) obviously could not be detected by these means. Evidently comparison of feedback from responses against "intentions" or "motor programs" necessarily implies that the subject detects his error because he knows what response he should have made.

A finding that a subject can detect errors without being able to specify the responses which he should have made would have eliminated model C, but could have been explained by either A or B without modification. The actual, contrary finding—that subjects know what response they should have made when they correct an error—could be explained by C without modification, but would require modification of either A or B. The most simple hypothesis which, without any modification, covers all the experimental results known to the author (Burns, 1965; Rabbitt, 1966*a*, 1966*b*, 1967; Rabbitt and Phillips, 1967), is thus that subjects only detect their motor errors, and that they do this by comparing feedback from a response against a motor programme which remains available after the response has been made.

Unfortunately it would indeed be naïve to suppose that humans detect all their errors in the same way in all possible perceptual-motor tasks. It is known that strategies of error-detection change with practice (Rabbitt, 1967). Different tasks impose different loads on different aspects of performance. Optimal strategies for detection of errors due to perceptual misidentifications (if any are possible!) may be expected to differ from good strategies in Experiments 1 and 2, where signals were very discriminable, and where it is probable that most errors were due to failure in the programming of responses. Nothing that we know about human behaviour suggests that people will rigidly confine themselves in all situations to a strategy which is optimal only for some cases. It is possible that there are many tasks in which subjects make errors which they cannot correct without feedback from the experimental equipment. For example, it is a common observation that many typing errors may occur because subjects move their hands up or down relative to a keyboard, thus making "correct" reaches to the wrong rows of keys. In this connection it is important to note that keyboard tasks more complicated than those used in Experiments 1 and 2 have shown much lower rates of error-detection. For example, Conrad and Longman (1965) have found that about 20 per cent. of errors were corrected during training on a chord keyboard, and that 23·3–34·4 per cent. of errors were corrected during typewriter keyboard training.

The necessary conclusion—that no single model is likely to describe all error-detection strategies used in all psycho-motor tasks—does not mean that we should abandon an attempt to rigorously define the various possibilities, and to select between them for particular experimental situations. It does, unfortunately, mean that we are unlikely to achieve an effective comprehensive description of these processes until much more evidence is available.

REFERENCES

Burns, J. T. (1965). *The Effect of Errors on Reaction Time in a Serial Reaction Task.* Unpub. Ph.D. thesis, Univ. of Michigan.

Conrad, R., and Longman, D. (1965). Standard typewriter versus chord keyboard—an experimental comparison. *Ergonomics*, **8**, 77–88.

Laming, D. R. J. (1968). *Information Theory of Choice-Reaction Times*. London: Academic Press.

Miller, G. A., Galanter, E., and Pribram, K. H. (1960). *Plans and the Structure of Behaviour.* New York: Holt.

RABBITT, P. M. A. (1966*a*). Errors and error-correction in choice-response tasks. *J. exp. Psychol.*, **71**, 264–72.

RABBITT, P. M. A. (1966*b*). Error correction time without external error signals. *Nature*, **212**, 438.

RABBITT, P. M. A. (1967). Time to detect errors as a function of factors affecting choice response time. In SANDERS, A. F. (Ed). *Symposium on Attention and Performance*, 131–42. Amsterdam: North Holland Publishing Co.

RABBITT, P. M. A., and PHILLIPS, S. (1967). Error-detection and correction latencies as a function of S–R compatibility. *Quart. J. exp. Psychol.*, **19**, 37-42.

Manuscript received 26th September, 1967.

SHORTER ARTICLES AND NOTES

EFFECTS OF AGE ON NOVELTY REACTIONS AND EXPLORATION IN RATS

BY

R. N. HUGHES*

From the Department of Psychology, The Queen's University of Belfast

Female rats of three ages (approximately 39, 160 and 316 days) were individually observed in an exploration box which consisted of a novel and a familiar half. By means of a behaviour-sampling technique it was recorded which half of the box the rat was in and what sort of activity it was engaged in at the end of each of 90 10-sec. periods. The number of 8 × 8 in. units entered was also recorded. All rats were seen in the novel half more often than in the familiar half and this novelty preference was greatest for the youngest rats. However, exploration and inter-unit locomotion scores tended to be lower with the younger animals. For all three groups these two scores showed a significant decrement during an observation period.

INTRODUCTION

In 1958 Broadhurst reported that ambulation in an open field was greater for rats about 100 days old than for 50 day old animals, but there was no significant difference between rats approximately 100 and 240 days old. However, Furchtgott, Wechkin and Dees (1961) later observed that open field locomotion decreased with age from 30 to 163 days old, but no further change occurred between 163 and 173 days. Candland and Campbell (1962) reported an increase in open field locomotion between rats 18, 23, 30, 44 and 54 days old but this age effect disappeared with rats older than 54 days. Williams, Carr and Peterson (1966) found that locomotor exploration in a symmetrical Y-maze increased with age in rats 27, 51 and 78 days old and then decreased between 78 and 116 days.

It is possible that the rather confused evidence of the above authors is related to the fact that their measure of exploratory behaviour was usually amount of locomotion. De Lorge and Bolles (1961) rightly feel that exploration cannot be adequately understood if it is tied to one response index. Therefore, it was decided to examine the effects of age on novelty preferences and several other measures of behaviour when each rat was allowed access to two environments differing in novelty.

METHODS

Subjects

The subjects were 36 female Wistar rats, all experimentally naïve. They were divided into three age-groups of 12 subjects each: (1) *Infant*, mean age = 39·3 days (range = 33–44 days); (2) *Adult*, mean age = 157·7 days (range = 155–162 days); (3) *Old*, mean age = 316·5 days (range = 296–354 days). Each subject had been weaned when 25 days old and caged with three other females of the same age. During the experiment all subjects lived in the experimental room and had *ad lib.* food and water up to and during testing.

Apparatus

The apparatus, which has been described in detail in a previous report (Hughes, 1965), consisted of four wooden exploration boxes, 16½ in. wide and 25 in. long. Each box was divided into six 8 × 8 × 8 in. units by means of incomplete walls ½ in. thick, 8 in. high and 2¾ in. wide. The walls had gaps in them 8 in. high and 2½ in. wide. The boxes could be divided into two halves by means of three guillotine slides, 8 in. high and 2½ in. wide, which could be later removed to give access to the six units—three "novel" and three "familiar." The interior of each box was painted mid-grey and the floor was always covered with fresh sawdust for each subject. Each half of a box was covered by a hinged wire-gauze lid from which a water dispenser nozzle was suspended in adjacent novel and familiar units at one end. Cubes of standard laboratory rat diet were also available in these two units.

---

* Now at the University of Canterbury, New Zealand.

Procedure

Each rat was confined to one half of a box with food and water for 12 or 24 hr. Following this, it was removed and placed in a cage while the guillotine slides separating the two halves were withdrawn thus enabling access to both halves. It was then put back into the same familiar half and 30 sec. later was continuously observed for 15 min.

By means of hand-counters the total number of 8 in. square units entered were recorded (*total unit entries*). Also, a behaviour-sampling technique was used to determine whether the rat was in the previously unavailable or novel half or not (*preference for novelty*), and what sort of activity it was engaged in at the instant a light from an electric timer flashed on and off, thereby signalling the end of a 10-sec. period. (There were 90 such observations for each rat.) Frequencies of the following forms of behaviour were thus recorded:

1. *exploration*—activities possessing familiarization properties by bringing a rat's receptors into closer contact with its external environment. The specific forms of behaviour consisted of sniffing the air or the floor and walls of the apparatus; rearing up on hind legs usually while sniffing; walking forwards, backwards or sideways within a unit or from one unit to another;
2. *grooming*—licking, scratching, rubbing or biting the body;
3. *eating*—eating a rat cube;
4. *freezing*—crouching or standing on all four legs entirely motionless. A *miscellaneous* category was provided for responses which could not be included in any of the other four categories.

Half of each age-group was confined to the familiar side of the apparatus for 12 hr. and the other half for 24 hr. The use of four boxes enabled the observation of four rats a day. Assignment of boxes and familiar halves to subjects was pre-determined in a non-systematic fashion so that (1) three rats from each age-group were tested in each box, and (2) the novel half was left for six rats and right for the other six of each group.

RESULTS

Separate analyses of variance were performed on all measures. To reduce the variances and normalize the distributions all proportional data (i.e. all except *total unit entries*) were transformed before analysis by the formula: angle = arc sin $\sqrt{\text{percentage}}$ (Snedecor, 1956; Winer, 1962).

Age had a significant effect on *total unit entries*, *preference for novelty*, *exploration* and *eating*. Except for one significant interaction on *total unit entries* ($F = 3{\cdot}43$, $d.f. = 2, 30$, $p < 0{\cdot}05$), time of confinement to the familiar half failed to have any effect. Mean scores on each measure and F ratios are displayed in Table I.

TABLE I

MEANS FOR EACH GROUP ON EACH MEASURE OF BEHAVIOUR AND RESULTS OF F TESTS

Measure	12 *hr.*	24 *hr.*	F *d.f.* 1, 30	*Infant*	*Adult*	*Old*	F *d.f.* 2, 30
Total unit entries	53·61	55·44	<1·00	48·33	61·50	53·75	3·74*
Preference for novelty	61·00	62·78	<1·00	71·08	58·75	55·83	6·76†
Exploration	63·94	66·28	<1·00	58·92	71·50	64·92	5·01*
Grooming	3·17	4·72	3·50	2·75	3·50	5·58	1·28
Eating	16·61	8·11	1·01	15·75	7·92	13·42	4·56*
Freezing	2·89	5·11	<1·00	5·75	4·08	2·17	2·72
Miscellaneous ..	3·39	5·78	<1·00	6·83	3·00	3·91	<1·00

* $p < 0{\cdot}05$ † $p < 0{\cdot}01$

The differences between individual means of measures significantly affected by age were tested for significance with t tests recommended by Lindquist (1953), using the mean squares for *errors* as estimates of variance and 30 degrees of freedom. The results were as follows:—

Total unit entries. *Adult* rats entered significantly more units than *Infant* rats ($t = 2{\cdot}73$, $p < 0{\cdot}02$), but there was no significant difference between *Adult* and *Old* rats

($t = 1\cdot61$), or between *Old* and *Infant* rats ($t = 1\cdot11$). (The significant interaction on this measure can be accounted for by increases in *Infant* and *Old* performances and a decrease in *Adult* performance between the 12 and 24 hr. confinement conditions.)

Preference for novelty. *Infants* were seen in the novel half at the end of significantly more 10-sec. periods than either *Adult* ($t + 2\cdot84, p < 0\cdot01$) or *Old* rats ($t + 3\cdot45, p < 0\cdot01$), but there was no significant difference between the latter two groups ($t = 0\cdot61$). As shown by binomial tests, significantly more than half of each age-group was seen in the novel half of the apparatus at the end of more than 50 per cent. of the 90 10-sec. periods (*Infants* $p < 0\cdot001$; *Adults* and *Old* rats $p = 0\cdot02$).

Exploration. *Adults* engaged in more exploratory activity than *Infants* ($t = 3\cdot16$, $p < 0\cdot01$) but did not differ significantly from *Old* rats ($t = 1\cdot67$). There was no significant difference between *Old* and *Infant* rats ($t = 1\cdot50$).

Eating. *Infants* ate food more frequently than *Adults* ($t = 3\cdot01, p < 0\cdot01$) but did not differ significantly from *Old* rats ($t = 1\cdot35$) who in turn did not differ significantly from *Adults* ($t = 1\cdot67$).

Inspection of Table I and consideration of the above t test results will reveal that the age effects are due mainly to significant differences between *Infant* and *Adult* rats.

Frequencies of each activity (except *miscellaneous*) over three successive 5 min. periods were determined for each age-group. Differences between the three periods were tested for significance by means of Friedman two-way analyses of variance. For all three age-groups there tended to be less exploratory and inter-unit locomotor activity near the end of the 15 min. interval than there was at the beginning (*exploration: Infant* and *Adult* $p < 0\cdot01$, *Old* $p < 0\cdot001$; *total unit entries*: *Infant* $p < 0\cdot01$, *Adult* and *Old* $p < 0\cdot001$). This also applied to *preference for novelty* scores for *Adult* ($p < 0\cdot05$). On the other hand, *Old* rats tended to show higher frequencies of *eating* near the end of the testing session ($p < 0\cdot02$). All other results indicated a general lack of change in frequency over time.

DISCUSSION

The *Adult* rats showed a greater number of *total unit entries* and a higher frequency of *exploration* than *Infant* rats, and yet *preference for novelty* was higher for *Infants*. This is surprising because if, within limits, amount of exploration is directly proportional to degree of novelty (Fowler, 1965), one might have expected the group that engaged in more *exploration* to show also the greater preference for the novel half of the apparatus. This large difference in *exploration* between *Infants* and *Adults* could be accounted for by the difference in *eating* between the two groups. In other words, by spending more time eating, the immature, rapidly "growing" *Infants* had less time available for exploring or conversely, by spending less time exploring they had more time available for eating. However, as inter-unit locomotion was recorded for both *exploration* and *total unit entries*, it is more likely that the *Adults'* higher frequency of the former was due to contamination with the latter. Therefore, it cannot necessarily be concluded that any "tendency to explore" was less marked with the *Infant* rats because of the possibility of *exploration* scores being influenced by either *eating* or *total unit entries*.

The locomotion scores (*total unit entries*) reflect a result similar to that reported by Broadhurst (1958), namely, an increase from very young to adult rats, but no significant difference between adults and older animals. Williams, Carr, and Peterson (1966) said that Broadhurst's result may have been due to the fact that his subjects came from an emotionally reactive strain, but as normal Wistars were used in the present experiment, this possibility seems slight. Williams *et al.* (1966) also found that on the first day of testing, their youngest group of rats (27 days old) spent more time in the arm of a Y-maze containing various small objects, than did any of their older groups. Assuming that this arm was the most novel of the three offered, it could be said that their youngest group showed the greatest preference for novelty. Even though the ages of the youngest groups differed in the two experiments, a similar conclusion can be drawn from the present experiment because the *Infant* rats spent the greatest amount of time in the novel half of the apparatus. This and the lack of differences in *freezing* between the three age-groups suggest that the initial timidity towards novelty shown by young chimpanzees (Welker, 1956) cannot be regarded as typical of female rats in the present experimental situation. Perhaps "fear of novelty" in the rat might depend more on developmental factors than has been previously recognized.

The significant decrements with time in locomotion (*total unit entries*) and *exploration* shown by all three age-groups have been observed in rats by several earlier workers (e.g. Bindra and Spinner, 1958; Montgomery, 1953).

It can be concluded from the results of this experiment that *Infant* rats showed the greatest preference for being in the novel half of the apparatus, but that in view of their *total unit entries* and *exploration* scores, *Adult* rats engaged in the most locomotor and exploratory behaviour. This latter conclusion is supported by the fact that the *Adult* rats were the only subjects showing a significant decrement in *novel preference* with time, a result suggesting a higher rate of novelty-habituation facilitated by exploratory behaviour. In view of the contamination of the *exploration* category with *total unit entries* scores, the greater amount of locomotion exhibited by the *Adult* rats might account for their higher frequency of *exploration*. However, it should not be assumed that the locomotion recorded in the *total unit entries* measure served absolutely no exploratory function at all.

REFERENCES

BINDRA, D., and SPINNER, N. (1958). Response to different degrees of novelty: the incidence of various activities. *J. exp. Anal. Behav.*, **1**, 341–50.

BROADHURST, P. L. (1958). Determinants of emotionality in the rat. II. Antecedent factors. *Brit. J. Psychol.*, **49**, 12–20.

CANDLAND, D. K., and CAMPBELL, B. A. (1962). Development of fear in the rat as measured by behavior in the open field. *J. comp. physiol. Psychol.*, **55**, 593–6.

DE LORGE, J., and BOLLES, R. C. (1961). Effects of food deprivation on exploratory behavior in a novel situation. *Psychol. Rep.*, **9**, 599–606.

FOWLER, H. (1965). *Curiosity and Exploratory Behavior.* New York: Macmillan.

FURCHTGOTT, E., WECHKIN, S., and DEES, J. W. (1961). Open field exploration as a function of age. *J. comp. physiol. Psychol.*, **54**, 386–8.

HUGHES, R. N. (1965). Food deprivation and locomotor exploration in the white rat. *Anim. Behav.*, **13**, 30–2.

LINDQUIST, E. F. (1953). *Design and Analysis of Experiments in Psychology and Education.* Boston: Houghton Mifflin.

MONTGOMERY, K. C. (1953). Exploratory behavior as a function of 'similarity' of stimulus situations. *J. comp. physiol. Psychol.*, **46**, 129–33.

SNEDECOR, G. W. (1956). *Statistical Methods.* Iowa: Iowa State College Press.

WELKER, W. I. (1956). Effects of age and experience on play and exploration of young chimpanzees. *J. comp. physiol. Psychol.*, **49**, 223–6.

WILLIAMS, C. D., CARR, R. M., and PETERSON, H. W. (1966). Maze exploration in young rats of four ages. *J. genet. Psychol.*, **109**, 241–7.

WINER, B. J. (1962). *Statistical Principles in Experimental Design.* New York: McGraw-Hill.

Manuscript received 16th October, 1967.

DEPENDENCE OF THE OVERTRAINING EXTINCTION EFFECT ON ATTENTION TO RUNWAY CUES

BY
D. LIKELY* and S. B. SCHNITZER
From The University of New Brunswick, Canada

Six groups of rats were given 60 extinction trials in a runway following 30, 90 or 120 acquisition trials. For half the groups, a hurdle was interposed in the runway during both acquisition and extinction. For rats tested with the hurdle, the 30-trial group extinguished more readily than the other two groups. With the hurdle absent, overtrained rats extinguished more readily. This was interpreted as supporting the hypothesis that overtraining transfers control of responding from external to internal cues, and that internally controlled running is relatively more susceptible to extinction.

INTRODUCTION

Conflicting results have been presented concerning the effects of overtraining on resistance to extinction. Some investigators (Bacon, 1962; Finger, 1942*a*; Ison, 1962; Madison, 1964; North and Stimmel, 1960; Seigel and Wagner, 1963) have reported that overtraining facilitates extinction. This has been termed the overtraining extinction effect (OEE). Others (Hill and Spear, 1963; House, 1963; Lawrence and Miller, 1947; Wilson, 1964) have found that overtraining increases the number of trials to extinction. A third group of studies (Bacon, 1965; Birch, Allison, and House, 1963; Lawrence and Festinger, 1962; Theios, 1962) presents findings showing that overtraining has no statistically significant effect on resistance to extinction. All of these experimenters obtained their data using a 100 per cent. reinforced, non-discriminative running response with rat subjects.

One suggestion (Mackintosh, 1965; Sperling, 1965) that may resolve the conflict is that facilitation occurs only in situations where training leads to a change in control of the response. Running may initially depend on external cues, but with increased amounts of training, control becomes internal, and the response is then relatively independent of exteroceptive stimuli. Further, Mackintosh (1965) and Sperling (1965) both hypothesize that internally controlled responses may be more susceptible to extinction. Thus, OEE would appear only when a comparison is made of the extinction scores for a highly trained, internally controlled response group with the scores of a less thoroughly trained group still responding in terms of environmental cues.

Transfer from exteroceptive to interoceptive control may be delayed or inhibited by increasing continued attention to external cues. This might be accomplished by increasing the distinctiveness of external stimuli, or by physically preventing subjects from developing the smooth and rapid running sequence typical of overtrained animals, or both.

Three levels of training were combined factorially with the presence or absence of a black hurdle interposed just before the goal-box of a white runway. It was predicted that in the hurdle condition overtraining would increase resistance to extinction, but in the non-hurdle condition, overtraining would facilitate extinction.

METHOD

Subjects

The subjects were 32 male and 28 female Sprague-Dawley albino rats, 68 to 74 days old at the start of training.

Apparatus

The runway was a straight, enclosed alley 6 in. wide, 12 in. high, and 50 in. long. Hand operated guillotine doors closed off a starting area 6 in. long and a goal area 8 in. long, both having the same width and height as the runway. The apparatus was painted a flat white, covered with ¼ in. mesh hardware cloth, and illuminated from above by three 7-watt bulbs placed at equal intervals 18 in. above the runway. The light was diffused and further reduced by a paper screen below the bulbs.

---

* Now at Dalhousie University, Halifax, Nova Scotia.

For three of the six groups, a black painted hurdle was placed immediately in front of the goal entrance. In order to reach the goal, a rat was forced to climb a 60° incline, cross the top of the hurdle 1·5 in. below the runway cover, and travel down a 45° incline to the goal-box threshold. The barrier was made of $\frac{1}{4}$ in. mesh hardware cloth bent over a black plywood form in the shape of an inverted T.

Procedure

Six groups of 10 rats were randomly chosen from the available subjects.

Two weeks prior to the start of training, all subjects were placed on a 23 hr. food deprivation schedule which was maintained throughout the experiment. All subjects were given 3 days of pretraining. For 10 min. periods on two successive pretraining days, subjects were individually placed in the start-box of the relevant (hurdle or non-hurdle) apparatus, with both start and goal-box doors open and no food present. On the third day, each rat was put directly into the goal-box and allowed to eat for 20 min., or until it had consumed 20 45-mg. Noyes pellets from a food cup.

Runway training began on the day following pretraining. The three hurdle and three non-hurdle groups were given 30, 90, or 120, continuously reinforced acquisition trials. Ten trials a day, each reinforced with one 45-mg. Noyes pellet were given until the desired number of trials was reached. The goal door was closed when all four of the subject's feet had crossed the threshold, and the time from opening of the start-box to closing of the goal-box door was recorded with a stopwatch. After entering the goal area, the rat was confined for 30 sec., and then immediately returned to the start for the next trial.

On the day following the end of acquisition training, 10 non-reinforced trials a day were given for 6 days. Criterion for a completed response continued to be putting all four paws across the goal-box threshold. Subjects were detained for 30 sec. in the goal-box between trials. Again, times were recorded with a stopwatch, and any subject that failed to complete a trial within 120 sec. was removed from the apparatus, and immediately placed in the start-box for the next trial.

RESULTS

Mean running times in acquisition were found to yield decreasing, negatively accelerated curves, for all groups. As a check on group differences in acquisition, running times were compared over groups and hurdle conditions at Trial 30, the last trial common to all groups. Presence of the hurdle significantly slowed running ($F = 6{\cdot}31$, $d.f. = 1, 54$; $p < 0{\cdot}05$). No difference was detected among acquisition level groups as such, however ($F = 0{\cdot}65$, $d.f. = 2, 54$, $p > 0{\cdot}05$), and the Groups × Hurdles interaction was non-significant ($F = 1{\cdot}55$, $d.f. = 2, 54$, $p > 0{\cdot}05$).

Two response criteria were employed as the extinction measures. For each subject, the total number of extinction trials which exceeded a response time of 60 sec., and those which exceeded 120 sec. were counted, and group means calculated (North and Stimmel, 1960). These means and their standard errors are presented in Table 1.

TABLE I

TRIALS EXCEEDING EXTINCTION CRITERIA: MEANS AND STANDARD ERRORS

Hurdle condition	*Acquisition trials N*	*60-sec. criterion*		*120-sec. criterion*	
		Means	*S.E.*	*Means*	*S.E.*
Non-hurdle	30	16·7	1·44	5·9	1·32
	90	24·6	2·36	10·9	1·92
	120	20·4	2·11	8·7	2·08
Hurdle	30	28·7	2·11	13·6	1·93
	90	18·8	1·60	8·1	1·44
	120	19·9	2·54	9·9	2·70

An analysis of variance for a 2 × 3 factorial design was performed on the data separately for the 60-sec. and for the 120-sec. criteria. For the 60-sec. criterion, neither the Trials effect ($F = 0{\cdot}76$, $d.f. = 2, 54$, $p > 0{\cdot}05$), nor the Hurdle condition ($F = 1{\cdot}25$,

$d.f. = 1$, 54, $p > 0{\cdot}05$) was significant. Similar results were obtained for the 120-sec. criterion (Trials: $F = 0{\cdot}03$, $d.f. = 2$, 54, $p > 0{\cdot}05$; Hurdles: $F = 1{\cdot}61$, $d.f. = 1$, 54, $p > 0{\cdot}05$). What is critical for the hypothesis is that, for both criteria, the Trials × Hurdles interaction was statistically significant (60-sec. criterion: $F = 9{\cdot}65$, $d.f. = 2$, 54, $p < 0{\cdot}01$; 120-sec. criterion: $F = 3{\cdot}64$, $d.f. = 2$, 54, $p < 0{\cdot}05$).

While the above analysis is useful in determining the predicted interaction, the nature of the analysis is such that the main effect of training on extinction is lost. Retardation of extinction by overtraining in the hurdle condition "cancels" facilitation of extinction by overtraining in the non-hurdle condition, as a main effect. Therefore, analyses were made treating the hurdle and non-hurdle conditions as if they were separate experiments. All groups were individually compared within hurdle conditions, using the 60-sec. criterion and a Scheffe *a posteriori* test. A further comparison was made between the 30-trial groups and the combined data for overtraining groups.

In the non-hurdle condition, the 30-trial vs. 90-trial difference was significant ($p < 0{\cdot}05$), and the combined test (30-trial vs. 90-trial plus 120-trial) approached significance ($p < 0{\cdot}10$). No other significant differences appeared in this condition. Similar analysis for the hurdle groups revealed that all differences were significant ($p < 0{\cdot}05$) with the exception of that between the two overtrained groups (90-trial vs. 120-trial). Direction of differences was reversed between the hurdle conditions.

Discussion

As predicted, overtraining was found to interact with the presence or absence of an interposed hurdle. Overtraining facilitated extinction when the runway was relatively homogeneous, but retarded extinction when the apparatus forced subjects to continue attending to environmental cues. These results are in accord with the hypothesis that overtraining serves to make running relatively independent of external cues, unless such cues are especially distinctive or interruptive, and that internally controlled running is more susceptible to extinction than is externally controlled running.

It was suggested that this hypothesis might help account for the discrepant results to be found in the literature. Several studies have employed conditions of 100 per cent. reinforced, non-discriminative training. Examination of the apparatus used does, in fact, lend credence to this suggestion. Of relevant experiments with an enclosed straight runway, eight demonstrate OEE (Clifford, 1964; Ison, 1962; Madison, 1964; North and Stimmel, 1960, two experiments; Seigel and Wagner, 1963; Theios and Brelsford, 1964; Wagner, 1963), and only one such study reports the opposite result (House, 1963). Others report no effect of overtraining (Theios, 1962; Wagner, 1961). This apparatus presumably encourages smooth interoceptive stimulus patterns; under the present hypothesis, transfer to internal response control should occur readily. Thus, Finger (1942*a*, *b*), carefully limiting distinctive cues, found evidence of OEE in an elevated apparatus. Lawrence and Miller (1947) suggested specifically that Finger's results were attributable to this lack of cues, and by deliberately differentiating the goal-box of their elevated apparatus by a well-lighted visual pattern, obtained reverse OEE. If Skinner box studies can be added, on the assumption that this apparatus is primarily a visual one, the almost unanimous finding that overtraining increases trials to extinction in that apparatus makes attention clearly relevant.

It might be pointed out that the use of a hurdle, both here and in Mackintosh's (1965) study, necessarily confounds increased attention with increased effort. Under at least some conditions, it appears that a more effortful running response will enhance resistance to extinction (Lawrence and Festinger, 1962). It can be hypothesized that one effect of effort is to increase the probability of sampling external cues. Thus, a high effort condition may contribute to inhibition of internal response control. We are continuing to study the effects of this and other variables on overtrained running responses.

References

Bacon, W. E. (1962). Partial-reinforcement extinction effect following different amounts of training. *J. comp. physiol. Psychol.*, **55**, 998–1003.

Bacon, W. E. (1965). Resistance to extinction following blocking of the instrumental response during acquisition. *J. exp. Psychol.*, **69**, 515–21.

Birch, D., Allison, J. K., and House, R. F. (1963). Extinction performance following discrimination training. *J. exp. Psychol.*, **65**, 148–55.

Clifford, T. (1964). Extinction following continuous reward and latent extinction. *J. exp. Psychol.*, **68**, 456–65.

FINGER, F. W. (1942*a*). The effect of varying conditions of reinforcement upon a simple running response. *J. exp. Psychol.*, **30**, 53–68.

FINGER, F. W. (1942*b*). Retention and subsequent extinction of a simple running response following varying conditions of reinforcement. *J. exp. Psychol.*, **31**, 102–33.

HILL, W. F., and SPEAR, N. E. (1963). Extinction in a runway as a function of acquisition level and reinforcement percentage. *J. exp. Psychol.*, **65**, 495–500.

HOUSE, R. F. (1963). *Resistance to Extinction as a Function of Sucrose Concentration and Number of Acquisition Trials.* Doctoral Dissert., Univ. of Michigan.

ISON, J. R. (1962). Experimental extinction as a function of number of reinforcements. *J. exp. Psychol.*, **64**, 314–7.

LAWRENCE, D. H., and FESTINGER, L. (1962). *Deterrents and Reinforcements.* Stanford Univ. Press.

LAWRENCE, D. H., and MILLER, N. E. (1947). A positive relationship between reinforcement and resistance to extinction produced by removing a source of confusion from a technique that had produced opposite results. *J. exp. Psychol.*, **37**, 494–509.

MACKINTOSH, N. J. (1965). Overtraining, transfer to proprioceptive control and position reversal. *Quart. J. exp. Psychol.*, **17**, 26–36.

MADISON, H. L. (1964). Experimental extinction as a function of number of reinforcements. *Psychol. Rep.*, **14**, 647–50.

NORTH, A. J., and STIMMEL, D. T. (1960). Extinction of an instrumental response following a large number of reinforcements. *Psychol. Rep.*, **6**, 227–34.

SEIGEL, S., and WAGNER, A. (1963). Extended acquisition training and resistance to extinction. *J. exp. Psychol.*, **66**, 308–10.

SPERLING, S. E. (1965). Reversal learning and resistance to extinction: A review of the rat literature. *Psychol. Bull.*, **63**, 281–97.

THEIOS, J. (1962). The partial reinforcement effect sustained through blocks of continuous reinforcement. *J. exp. Psychol.*, **64**, 1–6.

THEIOS, J., and BRELSFORD, J. (1964). Overlearning-extinction effect as an incentive phenomenon. *J. exp. Psychol.*, **67**, 463–7.

WAGNER, A. R. (1961). Effect of amount and percentage of reinforcement and number of acquisition trials on conditioning and extinction. *J. exp. Psychol.*, **62**, 234–42.

WAGNER, A. R. (1963). Overtraining and frustration. *Psychol. Rep.*, **13**, 717–8.

WILSON, J. J. (1964). Level of training and goal-box movements as parameters of the partial reinforcement effect. *J. comp. physiol. Psychol.*, **57**, 211–3.

Manuscript received 21st November, 1967.

INFANTILE HANDLING AND THE FACILITATION OF DISCRIMINATION AND REVERSAL LEARNING

BY

RODERICK WONG and JOHN L. JAMIESON

From the Department of Psychology, University of British Columbia

One group of rats was removed from their home cage and received daily handling from Days 3–21, while the control group of litter mates remained in the cage and did not receive any treatment. On Day 22 all rats were weaned and they were housed in individual cages until they were 60 days old. After two pretraining days, subjects were given daily blocks of one free and three forced trials in the T-maze and were rewarded with food after making the correct response. Following 20 days of training and testing on black-white discrimination, subjects were given 10 days of reversal training with four daily trials. Results indicate that the handled animals showed faster running and a greater number of correct choices than the control rats during both the acquisition and reversal learning phases of the experiment.

INTRODUCTION

A recent study by Wong (1966) has indicated that, in an appetitive learning situation, infantile handling did not facilitate the rats' performance in the T-maze. This result supports the previous findings of Spence and Maher (1962) and Du Preez (1964). Other experiments by Berstein (1957) and Goldman (1965) have yielded positive findings. A critical variable which may account for these conflicting results concerns the method of housing the animals during the post-weaning and test period. Whereas Goldman's (1965) animals were individually housed, those in Du Preez's (1964) experiment were housed four in each cage between weaning and testing. Similarly, Wong (1966) housed six rats in a large cage during the period between weaning and testing. It has been established by Ader (1965) that "group housing which effects a reduction in various behaviors purported to measure general emotionality in the rat, acts to decrease or eliminate differences between animals differentially manipulated during early life."

If it can be assumed that any facilitating effect of handling is due to some emotionality-reducing factor, the apparent discrepancy between the positive finding of Goldman (1965) and the negative results of Du Preez (1964) and Wong (1966) may be resolved quite simply. One of the implications of Ader's study is that the experimenter would be more likely to be successful in differentiating between the performance of handled and nonhandled animals in a learning task involving positive reward if they were housed in individual cages. This procedure was adopted in the present experiment.

METHOD

Twenty-seven female Wistar Albino rats bred in the laboratory were used. They were randomly assigned from their litters after birth to either the handled or the control condition, and were housed by litters until they were weaned at 21 days. Originally there were 14 rats in each group, but one of them in the control group died before the end of the experiment. The T-maze and the timing mechanisms were those described in Wong (1966).

The procedure used in the handling treatment was described in Wong (1966). On Day 22 all animals were weaned and were housed individually in 7 in. × 9½ in. × 7 in. cages on an *ad lib.* diet. When they were 60 days old they were placed on a diet of 15 gm. of food ration per day. They were habituated to this maintenance schedule for 8 days, and then given 2 days of pretraining in the T-maze. A sequence of free and forced trials was used in order to provide each subject with equal experience in the rewarded and nonrewarded goal-boxes. Four trials were given during each of the 2 days.

After pretraining subjects were given daily blocks of four trials on the discrimination task. A block consisted of one free trial followed by three forced trials. Half of the animals in each group were rewarded with kernels of Grapenuts weighing about 0·05 gm. for choosing the white goal-box, while the other half of each group received food for choosing the black goal-box.

The blocks were arranged so that each rat ran into the rewarded goal-box for two trials and into the non-rewarded box for the other two trials each day. The sequence of the trials after the first which was a free one was to force the animal into the side opposite to its initial choice for two consecutive trials. On Trial 4, they were forced into the side which they chose on Trial 1. The location of the rewarded goal-box was on the left during the first two trials and on the right during the last two trials.

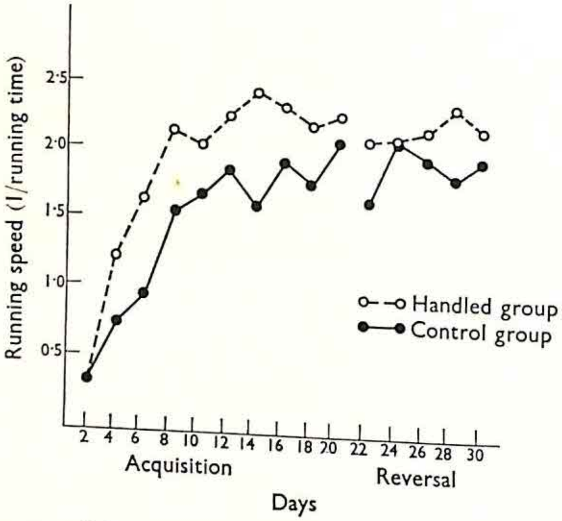
Following 80 acquisition trials which were given over a 20-day period the rats received discrimination-reversal training. The reversal series consisted of one free and three forced trials per day for a period of 10 days.

RESULTS

Acquisition

For analysis, all time measures were converted to speed scores (1/lat.). A comparison between the running speed scores of the handled and the nonhandled rats during acquisition revealed a significant difference ($F = 13{\cdot}96$, $d.f. = 1/25$, $p < 0{\cdot}005$). This is presented graphically in Figure 1 which shows a comparison of running speed scores of the handled and control animals during acquisition and reversal learning. It can be seen that during the initial days of acquisition the running speed scores of both groups were equally slow. Although both groups showed an increase in running speed with training ($F = 18{\cdot}34$, $d.f. = 19/475$, $p < 0{\cdot}001$), the handled group ran consistently faster than the nonhandled animals.

FIGURE 1



Running speed during acquisition and reversal learning as a function of infant treatment.

An analysis of the choice behaviour also reveals the superior performance of the handled animals. During the first 10 days of acquisition both groups were operating at a chance level in their correctness of choice; about 50 per cent. of the handled group's choices were correct while the control group made a score of 54 per cent. correct. However, during the last 5 days of acquisition (Days 16–20) the handled animals had attained a score of 93 per cent. correct in contrast to the 82 per cent. score obtained by the nonhandled group.

Reversal learning

An inspection of the reversal learning curves in Figure 1 also indicates that although the difference is not as pronounced as it was during acquisition, the handled group continued to run faster than the nonhandled group ($F = 5{\cdot}66$, $d.f. = 1/25$, $p < 0{\cdot}05$).

A comparison between the choice behaviour of the two groups during discrimination-reversal was analysed in the following way. For each subject the number of correct choices during the last 5 days during reversal learning was subtracted from the number of correct responses which it made during the last 5 days during acquisition. A *t*-test comparing the difference scores of animals in the handled and nonhandled groups was then computed, and their results indicate that the handled rats were able to learn the reversal shift faster than the control rats ($t = 1{\cdot}71$, $d.f. = 25$, $p < 0{\cdot}05$, one-tailed).

Discussion

It is quite apparent from these data that infantile handling does facilitate an animal's performance during the acquisition and the reversal of a discrimination response. It is believed that these results were obtained as a consequence of adopting the methodological suggestion proposed earlier. By housing rats in single cages during their postweaning period, we have minimized a factor which decreases differences between animals differentially treated during infancy.

It could be argued that the obtained results are attributable to differences in the motivational level due to body weight differences between the two groups. Hinde (1966) has pointed out that the weight at weaning of handled animals is greater than that of their nonhandled counterparts. Barnett (1963) has also indicated that handling facilitates the growth rate of the rat. From this it follows that the difference in growth rate is accompanied by an improvement in learning ability. However, a recent study by McMichael (1966) has shown that although handled rats were heavier than nonhandled ones at weaning, their weights were no longer significantly different at age 57 days. It is obvious that future studies on this handling and learning should consider and assess the effects of body weight loss of the groups.

This research was aided by grants to the first author from the National Research Council of Canada, APA-247 and the Medical Research Council of Canada, MA-2665.

References

Ader, R. (1965). Effects of early experience and differential housing on behavior and susceptibility to gastric erosions in the rat. *J. comp. physiol. Psychol.*, **60**, 233–8.

Barnett, S. A. (1963). *The Rat: A study in Behaviour*. London: Methuen.

Berstein, L. (1957). The effects of variations in handling upon learning and retention. *J. comp. physiol. Psychol.*, **50**, 162–7.

Du Preez, P. D. (1964). The persistence of some effects of handling in infancy on the behaviour of the adult rat. *Quart. J. exp. Psychol.*, **16**, 147–55.

Goldman, P. S. (1965). Conditioned emotionality in the rat as a function of stress in infancy. *Anim. Behav.*, **13**, 434–42.

Hinde, R. A. (1966). *Animal Behaviour: A Synthesis of Ethology and Comparative Psychology*. New York: McGraw-Hill.

McMichael, R. E. (1966). Early-experience effects as a function of infant treatment and other experimental conditions. *J. comp. physiol. Psychol.*, **62**, 433–6.

Spence, J. T., and Maher, B. A. (1962). Handling and noxious stimulation of the albino rat: II. Effects on subsequent performance in a learning situation. *J. comp. physiol. Psychol.*, **55**, 252–5.

Wong, R. (1966). Infantile handling and performance in the T-maze. *Psychon. Sci.*, **5**, 203–24.

Manuscript received 15th October, 1967.

CHANCE RESULTS IN UTROCULAR DISCRIMINATION

BY

W. B. TEMPLETON and F. A. GREEN

From Department of Psychology, University of Durham

A review is made of studies on whether we know with which eye we are seeing. The almost total failure to achieve chance results is ascribed to lack of control of convergence and it is shown that such control completely eliminates discrimination.

INTRODUCTION

In everyday situations it is easy, by closing one eye, or, less trivially, attending to the boundaries of the visual field, to decide which of one's eyes is receiving a particular stimulation. The classical problem then resolves into the almost tautologous question, familiar to students of E.S.P., of whether it is possible to discriminate when all known cues are eliminated. Nevertheless there remains the empirical problem of whether one can devise a situation in which a change in stimulation, which is itself discriminable, cannot be referred to the eye or eyes in which it occurs. Helmholtz (1925) believed that one could, and despite the almost total failure of attempts to do so, his assertion has been widely accepted ever since.

These attempts have involved two types of approach to the problem of selectively altering the stimulation of one eye. One group (Heine, 1904; Brückner and von Brücke, 1902; Thelin and Altman, 1929; Pickersgill, 1961) allowed unrestricted direction of gaze, placed an occluder close in front of one eye and flashed a light. Results were mixed but some subjects at least, discriminated with consistent accuracy. However, lack of convergence control allows the operation of several possible cues: Smith (1945) suggested that in the oculomotor adjustment to fixate on the stimulus the muscles of the open eye lead those of the closed eye and that the resulting differential feedback could provide the basis for discrimination. Perhaps more plausibly we might assume that the resting point of convergence in these experiments might be in approximately the same direction though not at the same distance as the stimulus; in this case the perceived azimuth direction of the stimulus (mediated either by retinal local sign or by feedback from the initial stages of the fixation reflex) would bear a constant relationship to the eye which was stimulated. (It is to be noted that the utilization of cues of this sort is essentially a two-stage process—the stimulus must be discriminated as appearing sometimes on the left and sometimes on the right and then this difference must be correctly interpreted. In the case of perfect discrimination of the cue and a consistent interpretation, we would expect a subject to be always wrong or always right.)

Pickersgill's (1961) experiment, which was essentially a repeat of the earlier work, has been widely cited in support of Helmholtz's claim (Pickersgill and Jeeves, 1964; Hochberg, 1964; Bower, Goldsmith and Hochberg, 1964) but in fact her overall error rate was only half the chance rate. All that can be concluded is that the ideal case mentioned above does not occur frequently and, as the earlier workers had found, longer exposure lowers error rates slightly; this might be expected whatever the cues on which the discrimination was based.

The final study in this group is that of Barrett and Williamson (1965) who used a more sophisticated technique for occluding vision and again recorded an error rate about half the chance level. But their stimulus was a solid model of a car, so that it would seem fairly easy for even unsophisticated subjects to work out which view belonged to which eye.

The alternative technique to stimulating only one eye is to stimulate both eyes with identical stereograms and then alter one of the stereograms. Since the crucial stimulus change is produced distally one can more easily eliminate any suspicion of differential adaptation of the two eyes or of temperature changes or auditory cues associated with proximally located occluders, and one would appear to have overcome the problem of uncontrolled convergence, outlined above.

The only experiment using this technique seems to be that of Smith (1945) who used square-grid stereograms with fixation on one of the corresponding pairs of cross lines. The central cross line of each grid had a small hole through which a light could be presented briefly to one eye or the other. Results were again well above chance level. But even this experiment does not provide an unchallengeable refutation of Helmholtz's assertion. It was intended specifically to eliminate the eye movements and consequent differential feedback from the stimulated "leading" eye which Smith claimed had contaminated the classical experiments. Yet it is difficult to see how this can have been achieved since he explains his supplementary finding, that increasing the convergence required increased the error rate, by claiming that the strain of convergence interfered with the muscle cues. In this particular experiment fixation was on the central cross lines and if it had been adequate no eye movements should have occurred.

Pilot runs for the present study made clear the probable deficiencies in Smith's set-up. We fused identical stereograms each consisting of a single bold black plus cross with a small hole at the centre and presented on a white ground. Despite the fact that a single fused cross and hole was visible, when a light was flashed through one of the holes it typically appeared not in the position of the visible hole but some way along one of the arms of the cross. It might be possible to seek an explanation in terms of heterophoria, or latent squint, the tendency for one eye to change its position when a gross alteration of the stimulus to that eye (e.g. introduction of an eye cover or a Maddox-rod disc) disrupts binocular fixation. If a light appearing briefly in a hole were sufficient to produce this effect, and little is known of how gross the change must be, then, irrespective of which eye took up a new position, the after-image of the light in the stimulated eye would correspond spatially with some non-central point on the cross stimulus in the other eye. But one would still require complete suppression of the cross surrounding the light in the stimulated eye.

A more plausible explanation would seem to be that binocular fixation was not accurate in the first place. Thus the cross arrangement is seen to be a form of the fixation disparity test (Ogle, 1964) in which two identical peripheral patterns are binocularly fused (seen in single vision) but are shown to be stimulating non-corresponding points when the two centrally placed halves of a single vertical line, one half presented to each eye, are adjusted so as to appear aligned with one another. Thus, if fixation disparity occurred in our situation, the large crosses and their unilluminated holes would appear in single vision but when one hole was illuminated the compulsion to fusion of the holes would cease and each would be referred spatially to the eye which it stimulated or, more likely, the unilluminated hole would be suppressed and the other seen in a direction appropriate to its, non-central, retinal image. In either case flashes in the left and right eyes would appear in different azimuth positions. We suggest that Smith's square line grids correspond to the crosses of the present argument and that such discrimination as occurred resulted from inaccurate binocular fixation.

It should be noted that Smith discarded trials on which a displacement was obvious to the subject; however, displacements between flashes temporally separated by several seconds might well serve as cues for utrocular discrimination without themselves becoming "obvious" to the subject; our elimination of fixation disparity was based on a simultaneous test. Smith also balanced his lights for intensity by presenting a rapid series of alternating left and right eye stimuli which "appeared as a repeated flash from a single source"; this must mean that similar stimuli to left and right eyes within Panum's fusional area are, if alternated sufficiently rapidly, referred to the same primary visual direction, just as perceived depth can result from alternating views of stereograms of appropriate disparity.

Smith found performance much better with foveal than with peripheral stimulation; this would be expected if spatial separation of successive flashes was the crucial cue. Horizontally peripheral stimulation gave a higher error rate—approaching chance level—than vertically peripheral stimulation. This result could be explained if horizontal fixation disparity is more marked or less easily disregarded by the system (as informal pilot observations of our own suggest) and if, for a constant eccentricity, two targets are more easily discriminated if they lie on opposite sides of the primary vertical meridian: only vertically peripheral stimuli separated by horizontal disparity would appear right and left of the primary meridian.

An earlier set-up of our own involved a dark background with only two binocularly fused fixation points and a flash which occurred vertically above one of the fixation points; there was marked horizontal displacement of the flashes, though with such a radially homogeneous stimulus, there was probably a cyclotorsional element as well, as the eyes would be in their characteristic resting position when released from radial fusion demand, and cyclophoria is known to alter visual directions (Howard and Templeton, 1964).

Finally, Smith also obtained chance results when he used two lights symmetrically opposite one another with reference to the fixation point; on some trials the two lights stimulated the same eye, on others different eyes; the subject had to judge whether both eyes or only one eye had been stimulated. Here the two flashes could appear in any of 20 possible positions, which would increase the difficulty of detecting a displacement.

These final two results appear to provide the only support in the literature for Helmholtz's contention. Nevertheless the contention may be generally true if accurate binocular fixation can be ensured. We were enabled to achieve this by virtue of the fact that for each of our subjects there was a binocular fixation distance for which he was apparently orthophoric.

Method

The subject sat at a high table on which was mounted an optical bench carrying the eye piece (a base-out prism before each eye) of a Hamblin hand stereoscope so that the distance of the latter from the subject could be varied. At the distant end of the optical bench stood a cardboard screen three feet by two in which was inset a small metal plate. Two $\frac{1}{8}$ in. holes horizontally aligned at eye level and 6·5 cm. apart were drilled in the plate—cleaner holes than could be drilled in the card. A plain white, moderately illuminated wall 15 ft. behind the screen provided an identical background for each hole.

The subject looked through the stereoscope and adjusted its distance until the holes appeared to be in focus and fused. He then half-covered one of the holes with his finger. This was usually sufficient to break up the fusion and separate the complete hole from the half-covered one. Further minor adjustments of distance were made until a true fusion involving corresponding points was achieved. Whether or not this involved suppression of the half-covered hole the effect was striking—a single complete hole, part of it seen through a transparent crescent of finger. Repetition of the test showed this to be a stable condition. If utrocular discrimination is abolished one could conclude that this procedure eliminates fixation disparity; otherwise the question of its adequacy would remain open.

On each trial one of the holes was obscured from behind by a wooden spatula for about one second. From the subject's point of view the single fused hole appeared to be obscured. He was asked to judge in which eye the obscuring had occurred and was required to respond right or left on every trial. There were ten trials, five left and five right in random order. The subject fixated on the fused hole while waiting for the occlusion.

Fifteen males and 15 females served as subjects, almost all students of psychology in their early twenties.

Results

Each subject had a score indicating the number of correct responses. Table I shows the frequency of each score together with the expected frequency on the assumption that the responses follow the binomial distribution with $p = 0{\cdot}5$ (the obtained proportion was 0·507). Scores 0, 1, 2, and 3, and 7, 8, 9, and 10 were combined into a single category and the resulting χ^2 (4 *d.f.*) was 2·35, well below the 0·05 significance value of 9·49. Thus the data are compatible with a binomial model and we would seem to be justified in considering the 300 responses as a single sample of independent events and considering the relative

TABLE I

Expected and Obtained Frequencies of Scores

Score	0	1	2	3	4	5	6	7	8	9	10
Expected frequency	0	0·3	1·3	3·6	6·1	7·3	6·1	3·6	1·3	0·3	0
Obtained frequency	0	0	1	3	6	10	4	5	1	0	0

TABLE II

Relative Likelihoods (Two Decimal Places) of Various Population Values of p, Given the Sample Proportion 0·507

True value of p ..	0·400	0·425	0·450	0·475	0·500	0·525	0·550	0·575	0·600
Relative likelihood	0+	0·03	0·11	0·55	0·97	0·82	0·25	0·11	0+

likelihoods of various possible population values of p. The relative likelihood for any population value Po is the ratio of the probability of a binomial population with $P = Po$ generating a sample proportion of 0·507 to the probability of a population with $P = 0·507$ (the maximum likelihood estimate) generating the same sample proportion. Table II indicates that only values between 0·4 and 0·6 have non-negligible likelihoods and we conclude that in the absence of fixation disparity any utrocular discrimination which can reach the verbal response level is extremely slight.

This conclusion is further strengthened by the inability of the single subject we tried, to learn the discrimination in the course of 50 trials with knowledge of results following each trial. On the learning trials he had 24 correct responses, on the subsequent 50 test trials 25.

It would seem that without blinking and without a test stimulus large enough to show the shape of the visual field boundaries there are no cues available directly to mediate verbal or presumably motor responses of utrocular discrimination in humans. Fixation disparity, the various phorias, and interocular differences in the refractive system could all produce differences between left and right eye stimuli but the significance of these cues would have to be learnt, probably in the experimental situation since the task is not normally encountered in everyday life.

Pickersgill's (1961) warning that subjective reports that an effect is seen in a particular eye should not be used to probe the etiology of after-effects is strengthened by the present study. On the other hand much of the spatial behaviour of animals with little or no overlap of monocular fields must depend on information about which eye has been stimulated. And in humans there is little doubt that unambiguous stereopsis requires that the disparate inputs be referred to the correct eye. The precise point in the system where utrocular information ceases to be available is a question for further research.

REFERENCES

BARRETT, G. V., and WILLIAMSON, T. R. (1965). Judging with which eye one is viewing a three-dimensional scene. *Percept. mot. Skills*, **21**, 455–8.

BOWER, T., GOLDSMITH, W. M., and HOCHBERG, J. (1964). Stereodepth from after images. *Percept. mot. Skills*, **19**, 510.

BRÜCKNER, A., and VON BRÜCKE, E. Th. (1902). Nochmals zur Frage der Unterscheidbarkeit rechts-und linksaugiger Eindrücke. *Pflüg. Arch. ges. Physiol.*, **108**, 263.

HEINE, H. (1904). Zur Frage der Unterscheidbarkeit rects-und linksaugiger Gesichtseindrücke. *Pflüg. Arch. ges. Physiol.*, **101**, 67.

HELMHOLTZ, H. (1925). *Handbook of Physiological Optics*, Vol. III, SOUTHALL, J. P. C. (Ed.). The Optical Society of America.

HOCHBERG, J. (1964). Depth perception loss with monocular suppression: a problem in the explanation of stereopsis. *Science*, **145**, 1334–6.

HOWARD, I. P., and TEMPLETON, W. B. (1964). Visually induced eye torsion and tilt adaptation. *Vision Res.*, **4**, 433–7.

OGLE, K. N. (1964). Some considerations of fusion in binocular vision and in oculomotor imbalance. In BENDER, M. B. (Ed.) *The Oculomotor System*. New York: Harper and Row.

PICKERSGILL, M. J. (1961). On knowing with which eye one is seeing. *Quart. J. exp. Psychol.*, **13**, 168–72.

PICKERSGILL, M. J., and JEEVES, M. A. (1964). The origin of the after-effect of movement. *Quart. J. exp. Psychol.*, **16**, 90–103.

SMITH, S. (1945). Utrocular or "which eye" discrimination. *J. exp. Psychol.*, **35**, 1–14.

THELIN, E., and ALTMAN, E. R. (1929). Identification of monocular functions. *J. exp. Psychol.*, **1**, 79–87.

Manuscript received 7th September, 1967.

DISCRIMINATION OF MIRROR-IMAGES: CHOICE TIME ANALYSIS OF HUMAN ADULT PERFORMANCE

BY

ROBERT W. SEKULER and KEVIN HOULIHAN

From Department of Psychology, Northwestern University

For members of many species left-right mirror-image discriminations are more difficult than are up-down mirror-image discriminations. A previous study with children using simultaneous presentation of pairs of figures showed that the relative difficulty of such discriminations depended upon the way in which the figures were aligned. The present experiment with human adults uses choice times to measure the difficulty of the two kinds of mirror-image discriminations under two alignment conditions. While the relative difficulties are clearly affected by alignment, the experiment shows that overall, left-right discriminations are still the more difficult.

INTRODUCTION

Members of widely different species have great difficulty discriminating between figures which are mirror-images of one another (Sutherland, 1961). Even more striking is the consistent pattern of relative difficulties presented by different kinds of mirror-image discriminations. As Rudel and Teuber (1963) pointed out, for children, as well as rats and octopodes, discriminations between left-right mirror-images are far more difficult than discriminations between up-down mirror-images. For example, members of pairs like ⊐ ⊏ or ⊏ ⊐ are more likely to be confused than are members of pairs like ⊔ ⊓ or ⊔ ⊓. Working with first grade children Sekuler and Rosenblith (1964) found that the previously reported differential difficulty of left-right and up-down mirror-image discriminations could be reversed when stimuli to be compared were presented one above the other rather than side by side, as Rudel and Teuber had done.

Since the difficulties of mirror-image discriminations are of both theoretical concern for the understanding of stimulus analyzing mechanisms (e.g. Sutherland, 1961) and are of considerable practical interest for the study of reading (Gibson, 1965), we have extended the earlier research to the domain of adult human subjects. Specifically, we are interested in the extent to which the relative difficulties of left-right and up-down mirror-image discriminations may remain invariant over changes in the alignment with which the to-be-discriminated stimuli are presented.

Since it was expected that college students would perform nearly errorlessly, some measure of difficulty other than errors was needed. Choice times were recorded in addition to judgements proper, in the hope that these times would provide a good index of the relative difficulty of each discrimination (Sekuler, 1966).

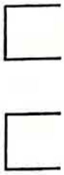
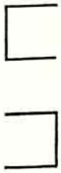
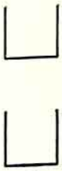
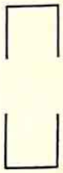
METHOD

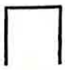
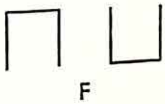
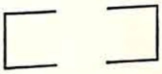
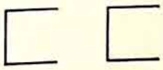
The stimuli used are shown in Table I. Each pair of figures was drafted in India ink and photographically imaged on high contrast 35 mm. slides. The slides were projected by a Carousel projector on to a Kodak rear projection screen mounted in one channel of a tachistoscope viewing box. A constant, low luminance background was provided by the other channel. The projected image was centred in a square aperture. Viewed over a distance of 57 cm., the largest dimensions of each figure subtended 10°40′ visual angle. The subject sat in front of the tachistoscope viewing box and on each trial was alerted (with the signal "now") when a stimulus would be presented. An electromechanical shutter on the slide projector was opened, causing the pattern to appear on the rear projection screen. The subject had been instructed to judge whether the members of each pair of stimulus figures had the same or different orientations. They were to indicate their "same"-"different" judgements by pressing a toggle switch in either of two directions. They were urged to make their decisions as quickly as possible with the additional constraint that they were to be "fairly certain" that their response would be correct.

Slides were presented at a rate of one each 5 sec. Each remained visible to the subject until he made his response, at which time the electromechanical shutter closed. Enough copies of the slides were arranged in the slide tray to permit each subject to make six judgements of each pair of figures. On half the trials, slides for stimulus types A, B, and H were shown after being rotated about a vertical axis, reversing the left-right characteristics

TABLE I

ERROR RATES, MEDIAN AND GEOMETRIC CHOICE TIMES FOR EACH STIMULUS TYPE

	A	B	C	D
				
Error rate	0·06	0·01	0·05	0·06
Geo. mean (msec.)	818	780	823	824
Median (msec.)	810	770	830	825

	E	F	G	H
				
Error rate	0·03	0·06	0·02	0·04
Geo. mean (msec.)	730	761	823	798
Median (msec.)	757	750	795	795

of those stimuli. On half the trials, slides for stimulus types C, E, and F were rotated about a horizontal axis, reversing their up-down characteristics. Slides were presented in six different random orders. For half the subjects, run in any given random order, a rightward switch throw designated a "same" response, a leftward throw a "different" response, while for the other half of the subjects the deisgnations were reversed. The interval between the opening of the shutter on the projector and the subject's response (choice time) was measured with a millisecond timer.

Subjects

The subjects were 24 undergraduate students from an introductory psychology course.

RESULTS

The median choice times for each subject and each of these eight sets of stimuli were found and entered in a Friedman analysis of variance by ranks. The Friedman test indicated that median times for the eight sets of stimuli were significantly different ($\chi_r = 20{\cdot}4$, $p < 0{\cdot}01$). Medians and geometric means of the choice times for each of the eight sets of stimuli are given in Table I. The Wilcoxon matched-pairs signed-ranks test was used to make those individual comparisons which are of experimental interest. The results of the Wilcoxon tests may be summarized as follows. When pairs of figures are arrayed laterally, our subjects took longer to respond to left-right mirror-image stimuli than to stimuli which are up-down mirror-images (type G vs. type F, $p < 0{\cdot}01$; type H vs. type E, $p < 0{\cdot}05$). When pairs of figures are arrayed longitudinally there was no statistically significant different between choice times for right-left and up-down mirror-image stimuli (type D vs. type B, $p > 0{\cdot}05$; type C vs. type A, $p > 0{\cdot}05$). Finally, when both types of mirror-image stimuli are aligned at right angles to their respective axes of reflection there is no significant differences between them. That is, subjects take as long to categorize right-left mirror-image stimuli aligned longitudinally as they do up-down mirror image stimuli aligned laterally (type G vs. type D, $p > 0{\cdot}05$; type H vs. type C, $p > 0{\cdot}05$).

As expected, the mean error rate (see Table I) was quite low in the experiment. The number of errors committed was so small that it is not possible to make meaningful comparisons among error rates for different stimuli.

Several recent studies have shown asymmetries in the times required for subjects to categorize stimuli as either "same" or "different" (Bindra, Williams and Wise, 1965; Nickerson, 1965). In order to determine whether a similar effect might obtain in our experimental situation each subject's choice times were compared for corresponding pairs of stimuli which yielded "same" and "different" responses. Sign tests were done on the times for pairs A and B, C and D, E and F, and G and H. None of these comparisons, between times to respond "same" and "different," approached statistical significance (all $p > 0{\cdot}25$).

DISCUSSION

In some conditions of this experiment subjects performed much like the elementary school children studied by Rudel and Teuber (1963) and Sekuler and Rosenblith (1964) in a similar task. These previous investigations showed that, when tested on laterally aligned figures, subjects made more errors in discriminating between left-right mirror-image stimuli than between up-down mirror-images. In the corresponding conditions of our experiment the college students did not make more errors but they did require a longer time to judge left-right than up-down mirror-images. The differential difficulty presented by these two classes of mirror-image stimuli is preserved despite a considerable difference, among experiments, in age and education of subjects as well as a change in the dependent variable used to measure difficulty of discrimination. The difference in choice times associated with the two kinds of mirror-image stimuli, in our experiment, was restricted only to the conditions of lateral alignment. Under conditions of longitudinal alignment there were no significant differences between left-right and up-down mirror-image choice times. This represents an apparent divergence from the earlier finding of Sekuler and Rosenblith (1964) that with longitudinal alignment up-down mirror-image confusions were more prevalent than were right-left confusions. This discrepancy between our findings and those of Sekuler and Rosenblith may not be as great as first seems. Sutherland (personal communication) pointed out that, taken as a whole, the data of Sekuler and Rosenblith show an asymmetrical difficulty of up-down and right-left mirror-images. With the lateral alignment of stimuli their subjects made only four confusions between right-left mirror-images and 25 between up-down mirror-images. Thus, while Sekuler and Rosenblith did find a modulation of difficulty of the two kinds of mirror image relationships as they went from one alignment to the other, even under longitudinal alignment the up-down discrimination was never as difficult as the left-right was under the lateral alignment. The present results like their earlier ones, then, show that while the differential difficulty of up-down and left-right mirror-image discriminations is modulated by the alignment in which the stimuli are shown, overall the left-right discrimination does seem more difficult then the up-down.

After completing our experiment we found that Butler (1964) has reported a related study using discrimination time to measure, among other things, the differential difficulty of left-right and up-down mirror-image stimuli. Using successive rather than simultaneous presentations of the figures to be discriminated, Butler found that times required to discriminate between left-right mirror-image stimuli were longer than those needed for up-down mirror-image stimuli. That finding is consistent with our own result that, considering both alignments used in our simultaneous discrimination situation, the right-left mirror-image judgements require more time than the up-down.

The fact that stimulus alignment clearly influences the difficulties of mirror-image discrimination means that one must be cautious in extrapolating from data collected with only one alignment to general theoretical propositions (Rudel and Teuber, 1963). If one is interested in extrapolating to tasks in which only one stimulus alignment is ordinarily used then data from research using a different stimulus alignment may be misleading. For example, in developing experimental analogues for reading it would seem most useful to insist that such analogues involve lateral alignment since in our culture reading materials are almost always so aligned (Gibson, Gibson, Pick and Osser, 1962).

In summary then our experiment shows that while the relative difficulties of left-right and up-down mirror images can be influenced by the alignment of the figures, in general the right-left discrimination is more difficult than the up-down. In this respect then human adults resemble the young of their own kind, as well as rats, and octopodes. Our experiment does not, of course, answer the two most interesting questions raised by ours and related work: (1) why is the right-left discrimination the more difficult? and (2) why is the difficulty of mirror-image discriminations influenced at all by the alignment of stimuli?

REFERENCES

BINDRA, D., WILLIAMS, J. A., and WISE, J. S. (1965). Judgments of sameness and difference: experiments on decision time. *Science*, **150**, 1625–7.

BUTLER, J. (1964). Visual discrimination of shape by humans. *Quart. J. exp. Psychol.*, **16**, 272–6.

GIBSON, E. J. (1965). Learning to read. *Science*, **148**, 1066–72.

GIBSON, E. J., GIBSON, J. J., PICK, A. D. and OSSER, J. (1962). A developmental study of the discrimination of letter-like forms. *J. comp. physiol. Psychol.*, **56**, 892–8.

NICKERSON, R. S. (1965). Response times for "same"-"different" judgments. *Percept. mot. Skills*, **20**, 15–8.

RUDEL, R., and TEUBER, H. L. (1963). Discrimination of direction of line in children. *J. comp. physiol. Psychol.*, **56**, 892–8.

SEKULER, R. W. (1966). Choice times and detection with visual backward masking. *Canad. J. Psychol.*, **20**, 34–42.

SEKULER, R. W., and ROSENBLITH, J. F. (1964). Discrimination of direction of line and the effect of stimulus alignment. *Psychonomic Science*, **1**, 143–4.

SUTHERLAND, N. S. (1961). *Shape Discrimination by Animals*. Experimental Psychology Society Monograph number 1.

Manuscript received 17th August, 1967.

SOME ANOMALIES IN THE STUDY OF EXPERIMENTAL PAIN

BY

DIANA R. HASLAM

From the Department of Psychology, University of Bristol

The experiment was designed in an attempt to throw some light on certain anomalies connected with the assessment of pain threshold by the Limiting Method. It is argued that an anchoring effect, similar to that described by Helson (1947), accounts for the low pain threshold values found when a small stimulus-interval is used.

INTRODUCTION

In a previous experiment (Haslam, 1965) it was found that when heat-pain threshold was assessed by the Limiting Method the size of the increment in successive values of the variable stimulus had a considerable effect upon the results. When the increment was as small as 8 mc./sec./cm.² pain threshold was significantly lower than when the variable stimulus was increased in steps twice or four times as large as this.

This finding was repeated in a further experiment (Haslam and Thomas, 1967), and in discussing the results the authors suggested that the factors which are operative when small stimulus-intervals are used are psychological, as opposed to statistical, in nature. In this connection the suggestion was made that subjects might perhaps tolerate only a limited number of trials; this was called the "tolerance limit." Hence, when the step-size is small, impatience might cause the subject to end the experiment by reporting pain at a low stimulus value. An alternative tentative explanation was that the suggestion of pain implicit in the instructions might induce particularly suggestible subjects to report pain after relatively few stimulus presentations. This would result in a low threshold value, especially when the step-size was small.

The experiment reported below was designed in an attempt to determine some of the psychological factors which operate in the perception of experimentally-induced pain. In order to do this it was decided to assess heat-pain threshold using two starting-points, one approximately half the value of the other, and two step-sizes, one a quarter of the value of the other. In this way it was hoped to throw some light on the tolerance limit hypothesis. In order to test the suggestibility hypothesis, a heat illusion test was used, as this type of suggestion was considered to be the most appropriate. There were five sessions and each subject served as his own control.

METHOD

Subjects

Fourteen subjects (seven male and seven female) were drawn from the student body and technical staff of the University. Their mean age was 20·7 years.

Apparatus and procedure

Pain threshold. Pain threshold was assessed by means of a modified Hardy-Wolff (1952) radiant-heat apparatus in which the heat from a 500-W. bulb was concentrated by a lens through an aperture 2 cm. in diameter. Mains power was supplied to the lamp through a Variac transformer and monitored by means of a voltmeter. The apparatus was calibrated by means of a Moll thermopile and millivoltmeter in terms of mc./sec./cm.², and the calibration was checked before each experimental session. Details of the procedure used are given elsewhere (Haslam, 1965). Two starting-points, 50 and 98 mc./sec./cm.², and two step-sizes, approximately 4 and 16 mc./sec./cm.², were used. These conditions will be referred to as the 50/4, 50/16, 98/4 and 98/16 conditions respectively.

Heat illusion test. The heat stimulus was again provided by the Hardy-Wolff radiant-heat apparatus. For this part of the experiment the stimulus energy was increased at a constant rate from zero until the subject reported a feeling of warmth. This was repeated six times, allowing a half-minute interval between successive trials. On the seventh trial, although the same procedure was simulated, no heat was actually applied. Further simulated trials were given until on three successive occasions the subject failed to report any sensation of warmth. A subject's suggestibility score was the number of times he reported warmth in the absence of a heat stimulus.

The four pain threshold assessments and the heat illusion test were carried out in a random order, and at least two days were allowed to elapse between each threshold assessment.

RESULTS

Pain threshold

Table I shows the mean pain threshold, range and number of stimulus presentations for the four conditions. Friedman's two-way analysis of variance (Siegel, 1956) showed that there was a significant difference in pain threshold as between these four conditions ($p <$ 0·001). A Wilcoxon matched-pairs signed-ranks test (Siegel, 1956) showed that, of the six possible combinations of starting-point and step-size, the only pair of pain threshold values which were not significantly different from each other were those for the 50/16 and 98/16 conditions. The 16 mc./sec./cm.² interval is the one which has previously been found to be the optimum interval in the sense that it gives rise to the most reliable pain threshold values (Haslam and Thomas, 1967).

TABLE I

MEAN PAIN THRESHOLD, RANGE AND NUMBER OF STIMULUS PRESENTATIONS FOR THE FOUR CONDITIONS REFERRED TO IN THE TEXT

Conditions	*Mean pain threshold in mc./sec./cm.²*	*Range*	*Mean number of stimulus presentations*
50/4	156	76–268	28
50/16	208	124–302	11
98/4	173	117–238	20
98/16	212	156–302	8

When the mean threshold values given in Table I were compared with 220 mc./sec./cm.², the value previously suggested as being the average heat-pain threshold for the whole population (Thomas, 1966; Haslam and Thomas, 1967), it was found that the values 156 and 173 mc./sec./cm.² were significantly different from 220 mc./sec./cm.² ($p <$ 0·001 and $p <$ 0·002 respectively) whereas 212 and 208 mc./sec./cm.² were not. For this purpose a *t*-test was used in which the null hypothesis was that the pain threshold for the whole population is 220 mc./sec./cm.²

Rank correlations as between threshold values for the various conditions can be seen in Table II. This shows that all the rho values were positive and significant at a probability level of 0·02 and better.

TABLE II

CORRELATIONS (RHO)

Conditions	*Rho value*	*Probability level*
98/16 and 50/16	+0·76	0·002
98/4 and 50/4	+0·88	0·001
98/16 and 50/4	+0·64	0·02
98/16 and 98/4	+0·75	0·002
50/4 and 50/16	+0·71	0·01
98/4 and 50/16	+0·87	0·001

Table I shows that the mean number of stimulus presentations ranges from 8 for the 98/16 condition to 28 for the 50/4 condition. When the conditions were paired according to interval-size, that is, 98/16 with 50/16, and 98/4 with 50/4, a Wilcoxon matched-pairs signed-ranks test showed a significant difference between 8 and 11 stimuli and also between 20 and 28 stimuli ($p <$ 0·01, two-tailed, in both cases).

The results for three female subjects are of interest because of the extremely low threshold values reported under the 50/4 condition. These can be seen in Table III.

TABLE III

PAIN THRESHOLD VALUES IN MC./SEC./CM.^{2} FOR THREE FEMALE SUBJECTS IN THE FOUR CONDITIONS

Subject	50/4	50/16	98/4	98/16
H	94	140	124	212
I	88	192	171	212
J	76	212	124	192

Heat illusion test

Only two of the fourteen subjects scored on the suggestibility test. One scored ten and the other scored one. Under the suggestibility hypothesis these two subjects, both of whom were female, might have been expected to have the lowest pain thresholds. This, however, was not so, and the suggestibility hypothesis was not confirmed.

DISCUSSION

The results confirm, once again, that interval-size has a significant influence upon pain threshold values, and they indicate that the starting-point, in combination with a very small step-size, also has a significant effect. Under the tolerance limit hypothesis pain threshold values would be expected to be determined by the number of stimulus presentations. This, however, does not appear to be the case. The mean pain threshold value of 173 mc./sec./cm.^{2} for the 98/4 condition was significantly lower than 208 mc./sec./cm.^{2} for the 50/16 condition. However, the value 173 mc./sec./cm.^{2} cannot be attributed to the tolerance limit for stimulus presentations having been reached because subjects tolerated a significantly greater number under the 50/4 condition. The tolerance limit hypothesis was thus not confirmed.

The finding that the mean pain threshold value for the 50/16 condition was not significantly different from that for the 98/16 condition supports the view that 16 mc./sec./cm.^{2} is the optimum interval for the assessment of pain threshold (Haslam and Thomas, 1967).

The possibility that "sensitization" occurs in small stimulus-interval conditions has been discussed and dismissed in an earlier paper for the reason that when pain threshold is assessed twice consecutively, it is unchanged the second time (Haslam, 1965). Further evidence against a sensitization effect is provided by the present finding that there were significantly more stimulus presentations in the 50/16 condition than in the 98/16 condition, yet the mean pain thresholds for these two conditions were not significantly different from each other.

Since neither the tolerance limit hypothesis nor the suggestibility hypothesis was confirmed, some other explanation of the results must be sought. The problem as to why small stimulus intervals should have the effect of significantly lowering the pain threshold is an interesting one, and a possible explanation is as follows. A series of low intensity stimuli alters a subject's frame of reference with the result that his "adaptation level" (Helson, 1947) is lowered. If the pain threshold is considered to be a form of adaptation level (AL), that is, the point which is intermediate between non-painful heat and fairly intense pain, then a series of low intensity stimuli can be thought of as exerting an anchoring effect on the AL similar to that found in many other cases of psychophysical judgement.

This explanation is illustrated by the results given for three subjects in Table III. When a small step-size was used in combination with a low starting-point, pain was reported at a stimulus intensity below the starting-point used in other conditions (98/4 and 98/16). In these latter conditions the starting-point stimulus (98 mc./sec./cm.^{2}) was described as "warm." In contradistinction to this, when a step-size of 16 mc./sec./cm.^{2} was used in combination with a low starting-point, pain threshold values were significantly higher. This lends credence to an anchoring effect, as also does the finding that the results for the 98/4 condition were intermediate between the results for the 50/4 and the other two conditions.

Some experiments carried out by Grindley (1936) on the variation of sensory thresholds (pressure, touch and pain) with the rate of increase of stimulus intensity are interesting in the light of the present results. He found that the touch threshold and the differential threshold for pressure increased as the rate of change of stimulus intensity decreased, whereas the threshold for pain showed no significant alteration. Grindley points out the danger of speaking of a sensory threshold "as though it were a fixed quantity, defined sufficiently by referring to its sensory mode and its place of origin," and explains his results in terms of sensory adaptation. Unlike the present experiment, however, the stimuli he used were continuous, and whereas he found that a slow increase in stimulus strength gave rise to a high threshold, the results of the present experiment indicate the reverse of this, that is, a small step-size gave rise to a lower threshold than did a large step-size.

More in line with the present results are those of Smith (1966), who, using the finger algometer, found that the faster the pressure was applied, the higher the pain threshold. His view is that the finger behaves like a semi-elastic body in which, to produce a given degree of deformation, the faster the rate of pressure applied, the greater the pressure needed. Smith goes on to say that his results demonstrate that control of technique is essential if variables capable of producing marked alteration in pain threshold are to be eliminated from pressure pain experiments.

The finding that pain is reported at such excessively low values in the present experiment emphasizes, once again, the highly subjective nature of the pain experience. However, even if the position is adopted that absolute pain threshold cannot be measured, a case can be made, in view of the present results, that individual differences in pain threshold can be assessed. The high *rho* values found when correlating thresholds for the various conditions support this view.

In order to overcome some of the problems inherent in the study of experimental pain, an optimum interval for the assessment of pain threshold has been suggested by Haslam and Thomas (1967). Smith (personal communication) has pointed out that if a method for studying pain has been found to be reliable, then there is no reason, in spite of the highly subjective nature of the pain experience, why pain should not be studied experimentally. In any event, it is hoped that in studying pain by psychophysical methods, light may be thrown not only on the mechanism of pain perception, but also on the more purely psychological problems raised by the use of these methods.

REFERENCES

GRINDLEY, G. C. (1936). The variation of sensory thresholds with the rate of application of the stimulus. I. The differential threshold for pressure. II. Touch and pain. *Brit. J. Psychol.*, **27**, 86–95 and 189–95.

HARDY, J. D., WOLFF, H. G., and GOODELL, H. (1952). *Pain Senstaions and Reactions.* Baltimore: Williams and Wilkins.

HASLAM, D. R. (1965). The influence of stimulus scale-interval upon the assessment of pain threshold. *Quart. J. exp. Psychol.*, **17**, 65–8.

HASLAM, D. R., and THOMAS, E. A. C. (1967). An optimum interval in the assessment of pain threshold. *Quart. J. exp. Psychol.*, **19**, 54–8.

HELSON, H. (1947). Adaptation-level as a frame of reference for prediction of psychophysical data. *Amer. J. Psychol.*, **60**, 1–29.

SIEGEL, S. (1956). *Nonparametric Statistics for the Behavioural Sciences.* New York: McGraw-Hill.

SMITH, R. (1966). The use of pressure and chemical stimulation to investigate pain. *Supplement Proc. Roy. Soc. Med.*, **59**, 73–8.

THOMAS, E. A. C. (1966). On determining pain thresholds using the limiting method. *Quart. J. exp. Psychol.*, **18**, 270–2.

Manuscript received 22nd August, 1967.

GRADIENTS OF DISTORTION SEEN IN THE CONTEXT OF THE PONZO ILLUSION AND OTHER CONTOURS

BY

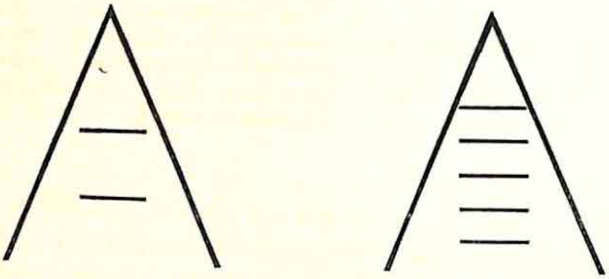
GERALD H. FISHER

From the Department of Psychology, University of Newcastle upon Tyne

It is suggested that the spatial distortion evident in the Ponzo figure is a special case of a more general illusion in which a gradient of attenuation appears within areas bounded by angular brackets. The magnitude of this gradient is measured in five lines seen against a number of angular contexts. A similar gradient appears also in the presence of single oblique lines. Accordingly, it is suggested that the distortions seen in the figures usually referred to as "the angle illusions" depend upon the presence of contours which do not necessarily define angles. The implications of these findings for certain existing theories which suggest that some illusions depend upon apparent-distortion of angular size and that they contain features usually associated with two-dimensional perspective projections of typical three-dimensional scenes are discussed.

INTRODUCTION

FIGURE 1



(*a*) The Ponzo illusion and (*b*) the general case of the distortion apparent in it.

Two identical horizontal lines placed within an inverted V-shaped bracket appear to differ appreciably in length. This illusory spatial distortion was first described by Ponzo (1912). The nature of the illusion is such that the upper line seems somewhat longer than the lower, this being seen clearly in Figure 1 (*a*). The distortion apparent in this figure is usually considered to be dependent upon the presence of the two inclined lines which define an angular spatial context against which the two horizontal lines are seen. Whilst considering the nature of the distortion evident in the Ponzo figure and in attempting to understand other illusions according to a common principle, it became apparent that the magnitude of the Ponzo illusion is influenced by the distance separating the two horizontal lines. Subsequently, it became clear that, rather than being a unique illusion, the Ponzo figure is a special case of the more general illusory spatial pattern illustrated in Figure 1 (*b*). In this figure a *gradient of apparent-distortion* of the lengths of the lines enclosed within the angular bracket may be seen clearly. The horizontal lines more nearly adjacent to the apex of the angular bracket appear successively longer than those which are isolated further from it. The purpose of this paper is to determine the nature and extent of the distortion-gradient evident in this new version of the Ponzo figure and to initiate consideration of the factors upon which its presence depends.

APPARATUS

An apparatus was devised using which five horizontal lines similar to those illustrated in Figure 1 (*b*), 5 mm. in width, could be varied in length symmetrically about their centres from 70 to 130 mm. with a position-reposition accuracy exceeding ±1/10 mm. These lines were seen as white bars against a matt-black background, 610 mm. square. The third line of the five was located centrally against the background, each further line being displaced by a distance of 50 mm. The angles and other figures used to provide the different spatial contexts were painted in white line upon sheets of a transparent plastic material. In all cases the lines defining these context-figures were 10 mm. in width. The plastic sheets were placed immediately in front of the background display and the bars could be seen clearly through them. The length of the line closest to the apex of the angle was set at 100 mm., this being used as the standard to which the lengths of the remaining four lines were adjusted by subjects. The method of adjustment was employed throughout and appropriate forms of instruction were given prior to carrying out each experiment. The three experiments described were undertaken individually in a randomly determined order by the same sample of 50 subjects.

EXPERIMENT I

Method

A sheet of transparent plastic painted upon which were two oblique lines, each 430 mm. long, inclined in such a way as to subtend an angle of 45 degrees, was placed before the background display. Being exactly square, this sheet could be placed in four orientations so as to display erect, inverted, left- and right-pointing V-shapes. For erect and inverted displays the background pattern of five lines was as described above, while for left- and right-pointing angular patterns the apparatus containing the lines and their associated adjustment mechanisms was rotated through 90 degrees, whereupon the lines assumed a vertical orientation. Each subject viewed all displays frontally being placed in such a position that the centre of each pattern was approximately 2 metres from the centre of his head. He was required to adjust each of the four variable lines, the initial lengths of which were randomised across the whole range of their possible adjustment, until all five lines appeared to be identical in length. The four directions of the V-shaped context were presented to subjects in a randomly determined order, this being changed each time the experiment was carried out. A control situation, omitting any context but inserting a plain plastic sheet, was included within each series. Each participant subject made one set of adjustments in each condition, thus making five series of responses in all.

Results

TABLE I

ESTIMATES OF THE MAGNITUDES OF DISTORTION INDUCED BY AN ANGULAR CONTEXT IN FOUR LINES JUDGED AGAINST A STANDARD LENGTH OF 100 MM.

	45° Angle															
	Upright		*Inverted*		*Left*		*Right*		*Angle omitted*		*Upright + Inverted*		*Left + Right*			
Line No.	$\bar{x}$	$\delta\bar{x}$	$\bar{x}$	$\delta\bar{x}$	$\bar{x}$	$\delta\bar{x}$	$\bar{x}$	$\delta\bar{x}$	$\bar{x}$	$\delta\bar{x}$	$\bar{x}$	$\delta\bar{x}$	$\bar{x}$	$\delta\bar{x}$	t	p
2	4·9	0·38	4·8	0·34	3·6	0·33	3·5	0·30	−0·15	0·18	4·9	0·25	3·6	0·31	3·26	<0·01
3	6·3	0·43	6·1	0·50	5·0	0·44	4·6	0·36	+0·23	0·20	6·2	0·33	4·8	0·40	2·70	<0·01
4	7·6	0·53	6·9	0·57	5·9	0·58	5·5	0·41	+0·48	0·22	7·3	0·39	5·7	0·47	2·62	≈0·01
5	9·5	0·64	8·5	0·70	6·7	0·61	6·0	0·46	+0·83	0·30	9·0	0·46	6·4	0·54	3·66	<0·01

The results obtained from this experiment are entered in Table I. They are appreciated most readily if the original horizontal lines are understood as being numbered 1 through 5, the number 1 being allocated to the standard length of 100 mm., this being the line closest to the apex of the angle irrespective of the orientation of the overall configuration. The data entered in rows refer to the mean extents of distortion ($\bar{X}$) apparent in each of the remaining four lines numbered 2 through 5. Alongside these are entered the standard errors of estimate ($\delta\bar{x}$) associated with them. All results were expressed in units of 1 mm. in excess of the standard length to avoid unwieldy computation. Reference to these data reveals that the distortion of apparent-length appears progressively throughout each set of lines seen within the framework of the angle. In a statistical sense the extent of distortion according to each orientation of the angle differs significantly from the control situation in which the angle was omitted. The trend evident in each set of data referring to the different angular orientations is significant also. These samples of data cannot be considered to have been drawn from the same population of responses, however, since statistically significant differences appear between the results referring to similar relative locations of lines according to the different orientations of the angle. The series of distortions apparent in relation to left- and right-facing angles do not differ, nor do those referring to upward- and downward-facing angles, but the former series differ significantly from the latter. Accordingly, the appropriate sets of data have been combined as is shown in the last two columns of Table I where further analysis confirms them to differ significantly. These entries provide the best estimates of the progressive magnitudes of the distortion-gradient with respect to angles presented in different orientations.

EXPERIMENT 2

Method

As shown in the first experiment, the magnitude of the illusory influence of the angular context of the Ponzo figure differs according to the relative positions of the two horizontal lines located within it. Although it has not previously been demonstrated, as far as we are aware, the distortion is influenced also by the angle subtended by the inclined lines which form the V-shaped bracket. To evaluate the influence of different contexts upon the distortion-gradient, oblique lines subtending angles of 45, 60, 75 and 90 degrees were prepared upon sheets of transparent plastic as before. These were again placed against the apparatus displaying the five horizontal lines, the conventional inverted orientation only being used for each different angular condition. Each member of the sample of 50 subjects viewed the displays selected in a randomly determined order interspersed with which was a condition in which no angle was presented. All subjects made one series of judgements in each situation.

Results

TABLE II

ESTIMATES OF THE INFLUENCE OF ANGULAR SIZE UPON THE DISTORTION-GRADIENT

	Angle								*Angle omitted*	
	45°		60°		75°		90°			
Line No.	$\overline{X}$	$\delta\overline{x}$	$\overline{X}$	$\delta\overline{x}$	$\overline{X}$	$\delta\overline{x}$	$\overline{X}$	$\delta\overline{x}$	$\overline{X}$	$\delta\overline{x}$
2	4·9	0·38	2·5	0·27	2·4	0·23	2·4	0·24	−0·15	0·18
3	6·3	0·43	4·0	0·37	3·7	0·62	3·7	0·33	+0·23	0·20
4	7·6	0·53	5·3	0·43	4·9	0·46	4·5	0·40	+0·48	0·22
5	9·5	0·64	6·5	0·55	6·0	0·55	5·6	0·50	+0·83	0·30

As before, the data have been entered in Table II in such a way as to refer to the horizontal lines numbered 2 through 5, the number 1 being understood as the standard length of 100 mm. with respect to which the apparent-lengths of the remaining four lines were judged. As in the first experiment the distortion-gradient is clearly apparent. Each induced distortion differs significantly from the control condition. More important perhaps is the marked difference in extent of the distortion-gradient as a function of the size of the angle subtended by the oblique lines. The distortion induced by the 45 degree angle exceeds that of all the others and it is clear that the magnitude of the gradient is influenced directly by the size of the angles defining the contexts within which the horizontal lines are seen.

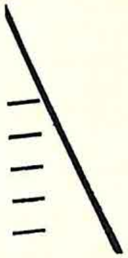
EXPERIMENT 3

The results of the first two experiments were such as to confirm that the Ponzo illusion is a special case of a distortion of visual space in which a gradient of attenuation appears within lines bounded by angular patterns. The nature of this gradient appears to depend upon *the presence of an angular form*, the influence of which is such as to induce distortion in other forms seen within its boundaries. It seems irrefutable that this is the principle according to which the distortion might be explained and gratuitous to question that it should be otherwise. Nevertheless, this is the problem considered in the remaining experiment.

Method

Experiment 2 was repeated with *single inclined lines* rather than complete angles, the left-hand arm of each angle and the left-hand halves of the horizontal lines being presented upon one occasion, the right-hand arm and lines on another. In this way the influence of each half of the angle upon the distortion-gradient was estimated. Lines inclined to left and to right at angles of 22·5, 30, 37·5 and 45 degrees to the vertical were used, presentation of the eight conditions being randomized along with an appropriate control situation in which the inclined line was omitted. One set of judgements was made by each subject in each condition. A typical stimulus situation used in this experiment is illustrated in Figure 2 in which the distortion-gradient induced by a single oblique line inclined at an angle of 22·5 degrees to the vertical may be seen.

FIGURE 2



The distortion-gradient induced by a single oblique line.

Results

TABLE III

ESTIMATES OF THE INFLUENCE OF INCLINED LINES UPON THE DISTORTION-GRADIENT

	Oblique line inclined to left (/)								*Oblique line inclined to right (\\)*								*Line omitted*	
	22·5°		30°		37·5°		45°		22·5°		30°		37·5°		45°			
Line No.	$\bar{X}$	$\delta\bar{x}$	$\bar{X}$	$\delta\bar{x}$	$\bar{X}$	$\delta\bar{x}$	$\bar{X}$	$\delta\bar{x}$	$\bar{X}$	$\delta\bar{x}$	$\bar{X}$	$\delta\bar{x}$	$\bar{X}$	$\delta\bar{x}$	$\bar{X}$	$\delta\bar{x}$	$\bar{X}$	$\delta\bar{x}$
2	2·2	0·15	1·4	0·13	1·1	0·13	1·1	0·13	2·3	0·21	1·4	0·15	1·4	0·15	1·3	0·13	−0·8	0·09
3	3·0	0·22	1·9	0·18	1·6	0·16	1·7	0·18	3·1	0·30	2·0	0·18	2·2	0·24	1·9	0·19	+0·12	0·10
4	3·5	0·28	2·6	0·24	2·3	0·23	2·2	0·22	3·5	0·33	2·4	0·23	2·6	0·29	2·2	0·24	+0·24	0·11
5	4·2	0·33	3·1	0·29	2·7	0·27	2·7	0·28	4·5	0·41	3·2	0·29	3·3	0·35	3·0	0·31	+0·42	0·15

The results of this experiment show that a marked distortion, similar in direction to that observed previously appears when the angle of the conventional Ponzo figure is replaced by a single oblique. Each of the mean values indicating the extent of this distortion differs significantly from zero and the trend of distortion is apparent in relation to each oblique line used.

More interestingly perhaps, the data from the experiment designed to consider the influence of different degrees of angular subtension and that from the present experiment have been reorganized in such a way as to allow direct comparisons to be made between them. In Table IV the mean values referring to the distortions induced by both angles and lines have been juxtaposed. The relevant comparisons are those in which the inclinations of the two lines together yield the appropriate angle. Appropriate values of Student's statistic t have been calculated in each case and the probabilities of occurrence of these on the relevant null-hypothesis have been determined. Without exception these may be interpreted as a clear indication that each angular distortion is the sum of the two individual distortions induced by the presence of the single oblique lines.

TABLE IV

COMPARISON OF THE EXTENTS OF DISTORTION BETWEEN ANGLES AND INCLINED LINES

Line No.	*Context*				t	p	*Context*				t	p
	45° Angle		*2 × 22·5° Lines*				*°60 Angle*		*2 × 30° Lines*			
	$\bar{X}$	$\delta\bar{x}$	$\bar{X}$	$\delta\bar{x}$			$\bar{X}$	$\delta\bar{x}$	$\bar{X}$	$\delta\bar{x}$		
2	4·9	0·38	4·5	0·37	0·85	>0·25	2·5	0·27	2·8	0·28	0·52	>0·25
3	6·3	0·43	6·1	0·53	0·29	>0·25	4·0	0·37	3·9	0·35	0·20	>0·25
4	7·6	0·53	7·0	0·60	0·81	>0·25	5·3	0·43	5·0	0·46	0·48	>0·25
5	9·5	0·64	8·7	0·71	0·94	>0·25	6·5	0·55	6·3	0·58	0·25	>0·25

Line No.	*Context*				t	p	*Context*				t	p
	75° Angle		*2 × 37·5° Lines*				*90° Angle*		*2 × 45° Lines*			
	$\bar{X}$	$\delta\bar{x}$	$\bar{X}$	$\delta\bar{x}$			$\bar{X}$	$\delta\bar{x}$	$\bar{X}$	$\delta\bar{x}$		
2	2·4	0·23	2·4	0·28	0·14	>0·25	2·4	0·24	2·4	0·46	0·09	>0·25
3	3·7	0·62	3·8	0·40	0·14	>0·25	3·7	0·33	3·6	0·37	0·20	>0·25
4	4·9	0·46	4·9	0·53	0	>0·25	4·5	0·40	4·4	0·46	0·25	>0·25
5	6·0	0·55	6·0	0·60	0	>0·25	5·6	0·50	5·7	0·58	0·07	>0·25

DISCUSSION

Descriptions of the nature of the distortion which characterizes the Ponzo illusion usually emphasize the influence of the angular context within which two lines, or other figures, of the same physical dimensions are seen. Previous studies of this illusion have been confined exclusively to evaluation of the relative distortions apparent in two such enclosed figures. No investigation has been reported, hitherto, in which the *gradient of distortion* seen throughout an angular figure has been demonstrated, or evaluated. The necessary and sufficient condition for the appearance of the Ponzo illusion is usually considered to be the presence of two oblique lines defining an angular context within the framework of which two otherwise similar spatial forms are seen. Many theories intended to explain illusions have been developed in terms of apparent-distortion of angular subtension. These have merely removed the problem a stage, however, since none has addressed itself directly to the question of why angles themselves should appear distorted in the first place. This remains an outstanding omission in our understanding of a fundamental feature of perceptual organization. The fact that a distortion, similar in nature and identical in extent, remains in the presence of the individual component lines of angular forms appears to point directly to a solution of this problem.

Many existing theories intended to explain the so-called *"angle illusions"* appeal implicitly, or explicitly, to the nature and direction of the distortion seen in the Ponzo figure. This illusion resembles a two-dimensional projection of a typical three-dimensional scene, increasing distance being indicated by convergence. Accordingly, the principles of perspective have been invoked widely in the attempt to explain the distortion seen in it, an early suggestion of this kind having been made by Thiéry (1896). More recently attempts have been made to extend understanding of the influence of typical perspective features upon the appearances assumed by two-dimensional spatial patterns in arguments which suggest size-constancy to be instrumental in inducing certain distortions. This argument was introduced by Tausch (1954) and taken up by von Holst (1957), Kristof (1961) and Gregory (1963). Some difficulties arising from these proposals have been discussed by Fisher (1968). The results of the present experimental study throw further doubt upon arguments of this kind since, although angular contexts may be strongly reminiscent of three-dimensional scenes, any depth features evoked by the presence of single oblique lines appear to be minimal.

The research reported in this paper was supported by Grant No. 70/GEN/9617 from the Ministry of Defence. The author is indebted to Mr. N. J. W. Horne of the University of Glasgow and Mr. J. J. Foster who assisted in carrying out these experiments.

REFERENCES

FISHER, G. H. (1968). Illusions and size-constancy. *Amer. J. Psychol.*, **81**, in press.

GREGORY, R. L. (1963). Distortion of space as inappropriate constancy scaling. *Nature*, **199**, 678–80.

VON HOLST, E. (1957). Aktive Leistungen der menschlichen Gesichtswahrnehmung. *Studium Generale*, **10**, 231–43.

KRISTOF, W. (1961). Uber die Einordnung geometrisch-optischen Täuschungen in die Gesetzmässigkeiten der visuellen Wahrnehmung. *Archiv. ges. Psychol.*, **113**, 1–48 and 127–50.

PONZO, M. (1912). Rapports de contraste angulaire et l'appréciation de grandeur des astres à l'horizon. *Arch. Ital. Biol.*, **58**, 327–9 (N.B. author refers to p. 239 of 1900 French edition of SANFORD, E. C. (1898). *A Course in Experimental Psychology*, Boston: Heath).

TAUSCH, R. (1954). Optische Täuschungen als artifizielle Effekte der Gestaltungsprozesse von Grössen- und Formenkonstanz in der natürlichen Raumwahrnehmung. *Psychol. Forsch.*, **24**, 299–348.

THIÉRY, A. (1896). Über geometrisch-optische Täuschungen. *Phil. Stud.*, **12**, 67–125.

Manuscript received 11*th November*, 1967.

BOOK REVIEWS

Timing in Temporal Tracking. By J. A. Michon. Soesterberg, Netherlands: Institute for Perception RVO-TNO, Netherlands. 1967. Pp. 127.

The author of this monograph uses the term "Temporal tracking" to describe the situation in which a human subject is required to reproduce by key tapping, either synchronously or successively, temporal sequences of auditory clicks. In a survey of earlier work on reproduction of time intervals he rightly points out the inadequacy of the available data for purposes of analysing sequential relations and in order to remedy this deficit he embarks on a series of experiments designed to elucidate the nature of the transfer function between input and output in key tapping tasks.

The first situation the author examines (Chapter III) is the ability of subjects to synchronise with, or to reproduce, isochronic (equal interval) sequences. He finds that the major part of the spontaneous variation may be considered as normally distributed random fluctuations in which the variance is proportional to the 1·5 power of the average response interval length. He points out that there is no simple stochastic process in which the variance increases at a rate greater than that of the mean and he resolves the problem neatly by suggesting that at each interval there may be some decay in storage of the representation of the immediately preceding interval. If the amount of decay is proportional to the time elapsed since the moment of storage by an amount reflecting a constant relative sensitivity (following Weber's law) this would give the relationship he obtains.

In Chapter IV the author goes on to examine responses to sequences in which the intervals are modulated in various ways, dealing with simple sine functions, composite sine functions, step functions and random input.

If one regards the human subject as an ideal linear predictor, the results from the sinusoidal conditions fit quite well with expectations, but in order to account for response characteristics to step function changes in input it is necessary to introduce further parameters; α, which is essentially constant for any given subject and β, which is a function of the average length of interval and which has the consequence of making the system nonlinear.

In Chapter V the author considers whether his results give any support to the notion of a time quantum or "perceptual moment" of relatively fixed duration. The results from the experiment on equal interval sequences show some evidence for periodicity in responding with a predominant value in the region of 80 msec. Another experiment in which subjects are required to adapt their tapping rates to slowly changing input rates shows stepwise changes in response rates which may be evidence for a further quantization or threshold effect. The author leaves open the question of whether or not these two effects are in any way related to the same basic process.

Chapter VI deals with some exploratory experiments in which key tapping is combined with other information processing tasks. The main conclusions are that if the temporal aspects of response are an integral part of the total task (as in piano playing) the information load of the non-temporal aspects of the task shows very little effect on temporal performance. On the other hand, if key tapping is used as a secondary task in conjunction with an unrelated primary task, then increasing information load in the primary task leads to greater variability in key tapping.

The final chapter provides an information processing model of the time sense, implemented as a FORTRAN program which should provide a useful basis for further investigations. Your reviewer feels some reserve about the enthusiasm shown by Michon in his introductory chapter for key tapping as the method *par excellence* for investigating all aspects of time perception. Temporal information is clearly used in different ways by the nervous system. Systems with quite different characteristics seem to be involved in, for example (*a*) the use of time differences in auditory localization (*b*) the perception of simultaneity or succession (*c*) the discrimination of true temporal order. However, Michon's conception of key tapping as a form of tracking and his analysis of input-output relationships in these terms is of undoubted value and the monograph deserves the serious consideration of all workers interested in the topic of the perception of time.

R. Davis.

Human Performance. By Paul M. Fitts and Michael I. Posner. Belmont, California: Brook/Cole Publishing Co. 1967. Pp. x + 162. 14*s*.

This paperback is one of a series called Basic Concepts in Psychology. The series is conceived as constituting a beginning textbook in psychology, with individual chapters appearing as separate volumes rather than under a single cover. Different volumes are written by different authors, drawn so far from the University of Michigan, and seventeen of these volumes have either appeared or will do so shortly. Teachers of psychology should find the series worth watching.

Outlined by the late Paul Fitts and completed by Michael Posner, this volume is a distinguished, and much needed, introduction to an important area of study which is described as follows. "The study of human performance, a branch of experimental psychology, analyses the processes involved in skilled performance, studies the development of skills, and attempts to identify factors which limit different aspects of performance. It seeks to analyse complex tasks into their simpler components and to establish quantitative estimates of man's abilities in each of the basic functions. In this way, it makes possible predictions about man's capability in performing complex skills." (p. 1.) This theme is filled out by seven main chapters which deal successively with: gross features of skilled performance, skill learning, effects of motivation, component processes and capacities, measurement of skills, perceptual-motor skills, and language skills. The writing is carefully lucid, a nice balance is held between general theoretical issues and particular experimental findings, and the bibliography and index are useful. The volume makes no pretence at being exhaustive but, considering its size and price, it presents a surprising wealth of material on human performance: more important, it welds this material into a coherent overiew of a kind which is not compactly available elsewhere. In brief, it succeeds in introducing an area of study which has progressed much in recent years and which ramifies widely into all other areas of psychological enquiry. However, the book's explicitly introductory character should not lull students into supposing that it can be read without effort, and entirely on its own. The reviewer has used the book with beginning students, and found it necessary to provide supplementary explanations, demonstrations, and laboratory experiments. The students found they had to read slowly and carefully, make frequent cross-references, and, to a large extent, comprehend the whole before they could comprehend the parts: they also found that several experiments were described so succintly that their essential features could not be appreciated. The experiences of these beginning students do not discredit the book—rather the reverse, they indicate the genuine intellectual challenge of its content. Since this book deserves many readers, particularly among undergraduates, it is as well for those readers to know that some background experience is necessary if they are to extract its full value. And its value is considerable because it provides undergraduate courses on skilled performance with exactly the kind of core text which has long been wanting.

I. M. L. Hunter.

Human Memory. By Jack A. Adams. New York: McGraw Hill. 1967. Pp. ix + 326. $9.50.

This book attempts to present the empirical data and the theories of contemporary psychological research in memory, and, as such, it is adequate, although not outstanding by any means. It covers verbal interference, serial and free recall, short-term and long-term memory, recall of verbal and motor responses, and recognition memory. It is extremely readable, assumes little previous knowledge of the subject, and is reasonably up to date.

My main criticism of the book is that Adams does not emphasize the need to make explicit assumptions about the underlying processes involved in memory. It is desirable to make explicit assumptions about how the underlying processes (e.g. acquisition, retention, retrieval and decision) interact to produce the observed response, before any conclusions are drawn from the observed response about one of the underlying processes. (Normally, experimenters desire to infer "strength" of an item in the retention process). For instance, Adams considers (p. 28) that there is no way of testing whether or not there is any change in strength when forgetting is observed in an interference situation. However, given a set of assumptions relating "strength" to observed response, it seems to me that it is possible in principle. Similarly, many of Adams's objections to Brown's decay theory disappear if one suggests that Brown was only referring to the strength of an item, and did not deny the existence of interference in other processes, like retrieval and decision.

Adams does consider the problem of measuring strength in Chapter 5, and advocates a latency index of strength, on the grounds that "latency is an established measure of strength" (p. 142). This is just not true. Adams's final chapter is interesting, as he considers omission and error-rejection behaviour in recall, but, instead of making the obvious inference that these phenomena reflect the action of a decision process, he postulates a "response recognition" mechanism. Not surprisingly, the most notable theorists omitted from this book are the only memory workers who do make the explicit assumptions mentioned above—the mathematical learning theorists and the Norman-Wickelgren school. Equally predictably, the main issue omitted from this book is one dear to both camps—whether the memory trace is continuous or discrete.

Apart from this major criticism, I have other minor criticisms of the book. For instance, I feel that Adams expends a disproportionate amount of space describing two recent experiments of his own on natural language mediation. Also, Adams gives Peterson and Peterson the credit for short-term memory research, whereas Broadbent (not mentioned in this connection) was just as influential. Furthermore, I thought that Adams should have considered objections to the STM/LTM duality, in view of the small amount of psychological evidence in its favour. Also, he speculates that recognition and recall involve different traces, and ignores the possibility that they might involve different retrieval mechanisms. Finally, Adams concludes that the directions memory research should take in the future should be towards studying mechanisms of forgetting, compartments of memory, and whether memory is permanent. It seems to me that, instead of taking these directions from the past, memory researchers in the future should first of all consider how to separate the processes involved in memory.

D. P. FARRINGTON.

Biological Foundations of Language. By Eric H. Lenneberg. With Appendices by Noam Chomsky and Otto Marx. New York and London: Wiley. 1967. Pp. xvi + 489. £6.

This is a real book, not a compilation. Perhaps it should be called a treatise. At all events, it brings together a great deal of material relevant to the biology of language and develops a firm and consistent point of view.

Basically, what Dr. Lenneberg has tried to do is to view language as a species-specific accomplishment, which while undeniably "learned," is at the same time—and in a very important sense—the outcome of maturation. Language, he believes, is based on underlying "cognitive functions" which are likewise specific to man, and which are thought to develop in the individual in much the same manner as the simpler innate behaviour patterns. These endogenous cognitive functions—viewed essentially as processes of categorization and the extraction of similarities—are brought to a state of "language readiness" at a particular stage of development (i.e. within a "critical period"), at which time the speech of adults seems to function as a releaser of the "developmental language synthesising process". Indeed the role of environment is regarded as little more than providing an appropriate context for maturation. If Dr Lenneberg is right, Dogberry may not after all have been so wide of the mark in supposing that such things as speech, reading and writing come by nature.

In short, Dr. Lenneberg is a nativist, a believer in language as a manifestation of "species-specific propensities" and as testifying to man's unique place in the evolutionary hierarchy. To this extent he finds common ground with Soviet psychologists such as Vigotsky and Luria, though, oddly enough, he seems totally unfamiliar with their work and viewpoint. Whether or not one agrees, at least his very intelligent argument is a much-needed counterpoise to the still powerful empiricist and neo-Behaviourist currents in present-day American psychology.

Dr. Lenneberg is perhaps at his best in discussing the problems of language-learning in deaf or otherwise handicapped children, in which his experience is clearly extensive. In this connection, he makes some telling criticisms of current remedial teaching methods and one must regret only his excessive politeness in voicing them.

This book is perceptive, thoughtful and on the whole well-written, apart from some unfortunate neologisms, such as "envision", "abolishment" and "evolvement." It should be extremely valuable as a text for advanced students and discussion classes and will doubtless stimulate a great deal of argument about the mechanisms of language learning and their neurological foundations. In this reviewer's opinion, it is a very impressive book indeed.

Dr. Chomsky contributes an Appendix on the formal nature of language, though its relation to what comes before strikes one as somewhat tenuous. There is also a second Appendix, by Dr. Otto Marx, on the history of the biological basis of language. In this, the author reminds us of the alleged experiment carried out by King Psanneticos of Egypt in the seventh century B.C., as related by Herodotus. It is said that he had two children raised by shepherds who were forbidden to speak to them for two years in order to see what language they would develop. Although Dr. Marx does not tell us so, what these two children are actually said to have produced was the Phrygian name for bread!

O. L. Zangwill.

The Genesis of Language: A Psycholinguistic Approach. Edited by Frank Smith and G. A. Miller. Cambridge, Massachusetts and London: The M.I.T. Press. Pp. xii + 400. £4.

This volume is the outcome of a conference on the development of language in children sponsored by the National Institute of Child Health and Human Development. With the exception of Hans Kalmus and Denis Fry, all the participants were American, and included D. I. Slubin, David McNeill, J. A. Fodor, Ira Hirsch, Eric Lenneberg, R. A. Chase, J. J. Jenkins and the late Ruth Weir. The proceedings follow the usual pattern of such conferences, though it is a relief to find that a good deal of the general discussion has been very strictly edited. Topics range from developmental psycholinguistics to a paper with the odd title of "Preparations for discussing Behaviourism with chimpanzee." There is an Appendix consisting of abstracts of some Soviet studies of language development in children.

So far as this reviewer is concerned, these papers are pretty specialized and several are rather dull. But it is good to see psychologists getting together with linguists and others in the hope of evolving really appropriate concepts and methods for the study of children's language.

O. L. Zangwill.

Acquisition du Langage et Développement de la Pensée. By H. Sinclair-de Zwart. Paris: Dunod. 1967. Pp. vi + 168. 23F.

The volume under review is one of the first major efforts to seek a rapprochment between the "mentalistic" psychology of Piaget and the "mentalistic" linguistics of Chomsky. As such it merits the careful attention of researchers in cognitive development and psycholinguistics. The investigation focusses on the acquisition of conservation and seriation skills in a group of Genevan children between the ages of four and eight. Supplementary results are given for a group of severely retarded children (aged 8 to 16), a group of senile dements, and a group of normal English-speaking children. In all cases, the data concerning the subjects' operational behaviour was compared with their ability to use relevant descriptive measure phrases appropriately. The penultimate chapter investigates the rôle of imagery and drawing skills in eliciting seriation behaviour.

Most of the general conclusions seem quite plausible. Thus the author finds support for such assertions as: (*a*) The development of both language and "thoughtful" behaviour are determined by underlying cognitive skills and hence are (on some level) functionally independent (due, presumably, to variables intervening between competence and "realization"). (*b*) The semantic component of linguistic skills serves to direct attention towards relevant task dimensions in concept-utilisation situations.

However, the "tough minded" validation of such simple and reasonable propositions as these raises very severe methodological problems. For example, the author seems to have assumed that the child's appropriate use of "parole" is sufficient to imply identity between the semantic structure of the child and the adult's "langue." Experimental evidence for this implication would greatly strengthen the author's case. (It should be mentioned that putative linguistic descriptions against which to check such experimentation are now becoming available, e.g. Ross and Bierwisch.) The volume is also flawed theoretically by a lack of concern for specifying the mechanisms responsible for transitions between cognitive levels. To isolate stages is clearly essential but surely developmental psychology should, at very least, explicate the notion of development. It would be unfair, though, to end on a carping note for Dr. Sinclair-de Zwart's studies are both an intrinsically valuable contribution to the experimental literature and a welcome bridge between the Piagetian tradition and recent Anglo-Saxon work on developmental psycholinguistics.

John C. Marshall.

Principles and Methods of Social Psychology. By E. P. Hollander. London and New York: Oxford University Press. 1967. Pp. xxiii + 520. 64s.

Current Perspectives in Social Psychology. Second Edition. Edited by E. P. Hollander and R. G. Hunt. London and New York: Oxford University Press. 1967. Pp. vii + 685. 40s. (paper).

Although literary twins do not usually appear in the annual demographic statistics, it is clear that the number of pairs of books born together of the same parents is on the increase. The fraternal twins from Hollander and Hunt are part of this increase. Both members of the pair are physically well-endowed and of very attractive appearance, and it may appear inhumane to recommend that they be separated from birth. This does however seem to be the best policy: they should be reared apart in quite separate environments.

It is apparent from the Preface of the two books that they are intended for different sets of readers, and it is difficult to envisage much overlap between these sets. *Principles and Methods* is avowedly introductory, whereas *Current Perspectives* is a book of contemporary readings from the primary research literature. Although there might be a need for each type of book, these volumes cannot readily be used together.

Consider first the text by Hollander himself. This covers well the field of social psychology and makes desirable excursions into neighbouring regions of history, sociology, personality theory and organization studies. The area dealt with is splendidly broad, but to tackle all this material in an introductory text is manifestly difficult. The emphasis is on stating the *principles* which have been evolved, and on supporting these with only selected evidence. Since many of the principles are rather low-level ones, the reader must inevitably experience some dissatisfaction.

Thus we have to make do with very sketchy treatments of a wide range of topics. In many places the accounts are just too simple—certainly not of a kind likely to lead an "introductory student" to pursue his studies further. Possibly the introductory student who plans not to get beyond an introduction may benefit, but hopefully there are not many of these. It is in any case a paradoxical aspect of introductory comments that they often make sense only *after* the reader is well past a beginning level. So although Hollander's comments are in the main sensible and to the point, the truly introductory reader will in fact probably not be aware of the point anyway.

The book undoubtedly improves as one moves through it. Quite naturally it is best when examining behaviour in small groups—a research area much enriched by the author's own work—and these sections may profitably be read by all students. But most other areas receive inadequate attention, and to understand and appreciate the tremendous quantity of research into attitudes, personality and developmental psychology one should preferably turn elsewhere.

Perhaps to *Current Perspectives* itself. This is the second edition of Hollander and Hunt's book of readings, and is at a quite different level from that of *Principles and Methods*. This second edition is larger than the 1963 version—containing 65 selections instead of 50—and it includes more than 20 contributions which have appeared since the first edition came out.

Current Perspectives in Social Psychology is an excellent book, ideally suited to courses in British universities. Its low price suggests that it may be recommended for student purchase, and the articles are judiciously selected to bring to life many of the points which are made in *Principles and Methods*. The book is organized into ten sections, each of which is preceded by a useful introduction by the editors. These introductions and the selections themselves make it clear how social psychology necessarily involves cultural and societal variables, how personality study is integral to the field, and how matters of national and international concern are becoming amenable to psychological research techniques.

The articles are drawn from a wide range of journals and from books of several orientations. Many of these are not readily accessible in this country, and it is this combination of varied material with a uniformly high standard which makes *Current Perspectives* a book to be recommended most strongly. Placing this twin in a stimulating foster-home should pay off handsomely.

PETER B. WARR.

The Use of Lateral Thinking. By Edward de Bono. London: Cape. 1967. Pp. 157. 18s.

Mr. de Bono believes that ordinary methods of education restrict and inhibit originality of thought (or "lateral thinking"), and his object is to suggest and to illustrate ways of breaking free from the bonds thus imposed. He does not discuss recent studies of "creativity," and his book is perhaps as near to Pelmanism as to academic psychology. However this may have its advantages as well as its drawbacks. Mr. de Bono is always lucid, and what he has to say may provoke thought as well as facilitating it.

ALAN GAULD.

"*Instinct*" *and* "*Intelligence*": *The Science of Behaviour in Animals and Man*. By S. A. Barnett. London: Macgibbon & Kee. 1967. Pp. xiv + 250. 45s.

This introductory book could be useful to undergraduates; although not as useful as the author's first book *A Study in Behaviour* which is more coherent and discursive. It is carefully written for a younger audience [e.g. on p. 40 the number 10^5 is translated as one hundred thousand], and, in most cases, technical terms are introduced smoothly. However, there are occasions when Barnett overreaches himself. Thus on p. 21 he quotes from von Neumann's last lecture-notes, on memory, "We are as ignorant of its nature and position as were the Greeks, who suspected the location of the mind in the diaphragm." Yet eight pages further on we see: "This is the cerebral cortex. It is here that the most complex patterns of behaviour are organized, and that the changes which we call 'learning' principally take place." This is not to impugn the quality of Barnett's thought, but it is the sort of issue (the relation between memory and adaptive change in behaviour as a result of experience) that merits a clearer exposition.

There are 16 good photographs, and many of the diagrams are very clear (although the drawings of brains are not as good as those in Romer). There are also many excellently concise sentences, e.g. "The male stickleback continues to guard the young when they have hatched. This entails suppressing any tendency to attack and eat them." However, there is a queer discussion of "nature and nurture" in which an example is taken from the adaptive aspect of bee-dancing rather than from a developmental problem. There are also many oversimplifications: ablation-studies and machine simulation in particular, and some sentences are misleading, e.g. "It might be expected that electrifying the brain would cause severe pain. . . ." It might not.

There is much good writing though: especially on herding and social interaction, insect sex-behaviour, and the brain (although he gives a cavalier treatment of the limbic system). The book concentrates on the natural history covered by "instinct" and "intelligence" rather than on the more precise and artificial topics which we would consider under "habituation," "rhythmicity," "S-R specificity," etc. That this is somewhat old-fashioned is indicated by the parentheses in the title. We are given the clue in the preface where the science of animal behaviour is considered to be ethology *tout court*, not experimental and comparative psychology. That may well be so; but then the book's subtitle *The Science of Behaviour in Animals and Man* is misleading.

Notwithstanding, investigation by experiment is a principal theme, and apparently is to be physiological in the main. This, and the process of natural selection are expounded as revealing the twin roots of behaviour. The emphasis on natural history is obvious if one compares the topics discussed under "exploration" in the two books. The second contains much less stimulating argument: the topic is not so obviously debatable. This is a general fault in the book.

Where Barnett does discuss a problem—crowding in animals, Stokes' work on blue-tit—then he is admirably clear and concise. The same is true when he mentions controls, e.g. in direction of migration in starlings, "nerve" section in Euplotes. There should be more of this; to stimulate lively thought, to make the reader more than a passive beholder of the wonderful diversity of nature, important though an admiring sense of wonder be.

NORMAN FREEMAN.

PUBLICATIONS RECEIVED

BOOKS

Tension in Medicine. Edited by Edmund Jacobson. Springfield, Illinois: Thomas. 1967. Pp. ix + 148. $7.75.

Instinkt Psyche Geltung. Bestimmungsfaktoren menschlichen Verhaltens. By Dieter Claessens. Köln and Oplade: Westdeutscher Verlag GmbH. 1967. Pp. 218. DM 29.50.

A History of Genetic Psychology: The First Science of Human Development. By Robert E. Grinder. London and New York: Wiley. 1967. Pp. xii + 247. 68s.

Individuality in Pain Suffering. By Asenath Petrie. Chicago and London: University of Chicago Press. 1968. Pp. xvii + 153. $5.0 45s.

Measuring Intelligence of Indian Children. By V. V. Kamat. London: Oxford University Press. 1968. Pp. xvi + 322. 45s.

The Nature of Human Intelligence. By J. P. Guilford. London and New York: McGraw-Hill. 1967. Pp. x + 538. 118s. 6d. $14.75.

Computers, System Science, and Evolving Society: The Challenge of Man-Machine Digital Systems. By Harold Sackman. London and New York: Wiley. 1967. Pp. xviii + 636. 116s.

Principles of Autonomic-Somatic Integrations: Physiological Basis and Psychological and Clinical Implications. By Ernst Gellhorn. London: Oxford University Press. (Minnesota University Press). 1967. Pp. xiii + 318. 80s.

PAPERBACKS

Some Thoughts about Teaching the Beginning Course in Psychology. By Edward L. Walker and Wilbert J. McKeachie. Belmont, California: Brooks/Cole. 1967. Pp. viii + 84. $1.75 text. $2.35 trade.

Sensory Processes. By Mathew Alpern, Merle Lawrence and David Wolsk. Belmont, California: Brooks/Cole. 1967. Pp. vii + 151. $1.75 text. $2.35 trade.

Quantification in Psychology. By William L. Hays. Belmont, California: Brooks/Cole. 1967. Pp. viii + 87. $1.75 text. $2.35 trade.

Noise and Society. By Michael Rodda. Edinburgh and London: Oliver & Boyd. Contemporary Science Paperbacks. 1967. Pp. viii + 113. 7s. 6d.

JOURNALS

Monitore Zoologico Italiano: Italian Journal of Zoology. Vol. 1, No. 1, 1967. Edited by L. Pardi. Quarterly. New Series. Most papers in English. Subscription $12.

Revista Peruana de Instruccion Programada No. 1. June 1967. Edited by Dr. Anibal Espino Rodriguez. Facultad de Letras y Educacion, Universidad Nacional de Trujillo, Peru.

Neuroscience Translations. Edited by Robert W. Doty. Number 1. 1967–68. Published by the Federation of American Societies for Experimental Biology for the National Institute of Mental Health. Translated material will cover all areas of basic research on the nervous system and will come primarily from the current Russian literature. Quarterly. $10.00 per year. (Subscription to first year *free* upon written request to Neuroscience Translations, 9650 Pockville Pike, Bethesda, Maryland 20014.)

TESTS

Family Relations Indicator. By John G. Howells and John R. Lickorish. Manual and test cards. Edinburgh: Oliver & Boyd. 1967. £7 7s. per set.

MONOGRAPHS

Mental Health Research Institute 11th Annual Report. 1966. University of Michigan. Pp. 93.

Fehlerquellen der Sog. ASW-Versuche. By Gösta Olof Lindholm. Helsinki: Finnish Academy of Science and Letters. Monograph Series B. Vol. 145, 2. 1967. Pp. 292. 26.0 F mk.

A Neural Systemic Theory of Emotion. By Philip L. Sawyer. Privately printed by Author, Bowen & Sawyer, 46 South Street, Auburn, N.Y. 13021. 1966. Pp. xii + 57. $2.95 Cloth. $1.75 Paper.

Psychologische Forschung

Empirische Untersuchungen von Grundproblemen der Psychologie

Basic Research in Psychology

With collaboration of numerous experts

In the interest of speedy publication this journal is issued at frequent intervals, according to the material received. As a rule 4 numbers constitute one volume. The price is DM 120,—; US $30.00 (plus postage) per volume.
Members of the "American Psychological Association" are granted a 15% price reduction.

Edited by

O.-J. Grüsser, Berlin
R. Held, Cambridge, Mass.
Th. Herrmann, Marburg.
H. Hörmann, Berlin
H. W. Leibowitz, University Park, Pa.
P. Leyhausen, Wuppertal-Elberfeld
D. Ploog, München
E. G. F. Sauer, Gainesville, Fla.
K.-H. Wewetzer, Giessen

Scope of the Journal

"Psychologische Forschung" invites contributions in general experimental psychology. Inasmuch as psychology is broadly concerned with problems of behavior and conscious processes, contributions from related behavior-oriented disciplines, such as zoology, ethology, physiology, psychiatry, sociology, and neurology are encouraged—provided they contribute to our understanding of basic psychological problems. Manuscripts submitted to "Psychologische Forschung" should be firmly based in experimentation and should, at the same time, explore the theoretical relevance of the data obtained. It is the sequence of theory stimulating experimentation and leading back to theory—applied to the domain of basic problems of psychology—which is especially welcome in "Psychologische Forschung". Exposition and style should not restrict the prospective readership of an article to the specialist in a particular discipline, but should be aimed at demonstrating the significance of the problem under consideration within the broader context of the behavioral sciences.

■ **Leaflet on request!**

SPRINGER-VERLAG
BERLIN · HEIDELBERG · NEW YORK

REVUE CANADIENNE
CANADIAN JOURNAL
DE PSYCHOLOGIE
OF PSYCHOLOGY

VOLUME 21, No. 6

DECEMBER 1967

Effects of handling at various ages on later open-field behavior: BARBARA A. DOTY & LARRY A. DOTY

Mediation or pseudomediation: a reply to Earhard and Mandler: DAVID L. HORTON

A pain-tolerance determination technique for evaluating analgesic drug efficacy: ALAN J. AUERBACH

Stimulus-response compatibility and parallel response selection: MARILYN C. SMITH

Stimulus determinants of the practice decrement of the Müller-Lyer illusion: ROBERT E. DEWAR

The perception of curvature: JOHN OGILVIE & EVA DAICAR

Mode de disparition des figures stabilisées sous différents angles visuels: B. CARDU & M. GILBERT

Preservation of fear measured by changes in rate of avoidance responding in dogs: MORRIE BAUM

Published for the Canadian Psychological Association by
UNIVERSITY OF TORONTO PRESS

THE CANADIAN JOURNAL OF PSYCHOLOGY is published bimonthly. Annual subscriptions, $10.00, single number $2.00.

Contributions. Original manuscripts and correspondence on editorial matters should be sent to:

The Editor, Canadian Journal of Psychology, McMaster University, Hamilton, Ontario, Canada.

Subscriptions to Journal. Orders and correspondence regarding subscriptions, changes of address, and purchase of back numbers should be sent to:

Canadian Psychological Association,
The Secretary-Treasurer,
Suite 210, 225 Lisgar Street,
Ottawa 4, Ontario, Canada.

QUARTERLY JOURNAL OF EXPERIMENTAL PSYCHOLOGY

ANNOUNCEMENT

Publication Arrangements

With effect from 1st January, 1969 the Quarterly Journal of Experimental Psychology will be published for the Experimental Psychology Society by Academic Press Inc. (London) Limited. Subscribers will receive renewal notices from the Academic Press. All transactions to take effect after 1st January, 1969 should be conducted with:

Academic Press Inc. (London) Limited,
Berkeley Square House,
Berkeley Square,
LONDON, W.1

or through booksellers as hitherto.

THE QUARTERLY JOURNAL OF
EXPERIMENTAL PSYCHOLOGY

Vol. XX 1968 Part III

SHADOWING IN BISENSORY MEMORY

BY

RHEA L. DORNBUSH*

From Douglass College, Rutgers—The State University

Shadowing (vocalization-at-presentation) was applied to a bisensory situation where different messages were simultaneously presented to the visual and auditory modalities. Three groups of subjects were employed: Group I shadowed the visual modality; Group II shadowed the auditory modality; Group III was a control, shadowing neither modality. Shadowing in the present experiment facilitated recall of the shadowed modality, particularly the visual modality, which is usually inferior to auditory recall.

It also became apparent that visual recall in an ordinary bisensory situation was minimal if not near an incidental level and that a true bisensory situation with equal division of attention between the two modalities employed does not exist.

INTRODUCTION

The present experiment is concerned with the effect of shadowing on bisensory simultaneous stimulation. Bisensory stimulation refers to the simultaneous presentation of different messages to two sensory modalities: in this case, the auditory and visual modalities. Shadowing or vocalization-at-presentation is the repeating out loud of the items on one channel or modality immediately as they occur.

Not much has been done with bisensory stimulation—particularly with verbal materials. However, it is known that under bisensory conditions more auditory material is retained than visual material. In addition, when rate of presentation is manipulated, the two modalities seem to operate independently of each other; that is, the visual and auditory curves are similar to those obtained when each modality is stimulated separately; the typical auditory decay function is obtained with decreasing or slowing rate whereas visual recall increases with decreasing rate.

The main interest of the present study was the effect that shadowing would have in altering visual results. Initially in an attempt to explain the limited visual recall, attention was given to the role of perceptual factors: e.g. processing time, read-out time, etc. However, the conditions of stimulus presentation were such as to preclude the use of such an explanation. That is, stimulus items were exposed for 0·5 sec., more than enough time for accurate processing.

One of the more plausible explanations of visual recall and also of comparisons between visual and auditory recall concerns itself with the nature of rehearsal. Both auditory and visual stimuli are thought to be transformed into auditory subvocalizations (Broadbent, 1963; Sperling, 1960; Wickelgren, 1965). Thus there is a translation factor that probably serves to retard visual recall particularly at the fastest rates. That is, auditory material when subvocalized is auditory—the subvocalization or rehearsal is in the same modality as the stimulus presentation. Visual material when subvocalized must be transformed or translated into auditory subvocalizations. It is possible that items are lost in the translation of the visual materials. However, as the interval between successive visual items increases there is an increased opportunity, which approaches the auditory level, to translate and rehearse the visual materials. It is also probable that auditory material is processed first so that when visual material is finally attended to and subvocalized, this may be done from a fading trace.

---

* Now at Department of Psychiatry, New York Medical College, New York 10029.

It was therefore thought that shadowing the visual items might provide a "simultaneous" translation of the visual material and possibly equate auditory and visual scores and result in similar functions with respect to rate of presentation.

In the present experiment a combination of variables previously studied in single stimulation and dichotic stimulation experiments were employed in a bisensory situation where the subject was instructed to shadow one of the two simultaneous messages. The variables of interest were: (1) over-all difference in the visual and auditory modalities; (2) rates of presentation of materials; (3) order of recall for a modality, either first or second; (4) type of material, number or letter, and combinations of materials in the two sensory channels; and (5) shadowing the message on one of the channels.

METHOD

Apparatus. Materials were presented on 16 mm. film (24 frames/sec.) with synchronous sound track by means of a Kalart projector. The exact auditory sequences intended for pairing with the appropriate visual sequences were recorded in a sound studio by the experimenter; precise control of synchrony was accomplished by splicing the auditory tape. That is, speech recordings were spliced in to the film so that each utterance of a stimulus item commenced within the first half frame of its associated visual stimulus (+21 msec.).

Visual stimuli were made by "Letraset" instant black letters exposed singly on white $8\frac{1}{2}$ in. × 11 in. cards which were then filmed and appropriately edited. The contrast ratio as displayed was approximately 80 per cent. The interstimulus and interlist intervals for visually presented materials consisted of clear leader which was slightly less bright than the stimulus sequence thus minimizing the negative after-image. The playback level of the tape recorder was so adjusted as to sound equal in loudness to the level of the sound track emanating from the projector's speakers. It seemed reasonable to assume that the intensities might not be the crucial variable so long as the material was clearly and easily discriminable.

List preparation and presentation. The first 10 letters of the alphabet, A through J, were employed as were the digits 0 through 9. These letters were chosen as pilot studies measuring tachistoscopic thresholds of the 26 letters of the alphabet indicated that each item was recognized with equal facility.

In the bisensory situation each list contained six items. Altogether 48 lists were required in order to produce 24 pairs of lists sampling all possible combinations of three rates of presentation, two orders of report, and four combinations of materials. The materials combinations were: (1) auditory-numbers, visual-numbers, (2) auditory-numbers, visual-letters, (3) auditory-letters, visual-letters, and (4) auditory-letters, visual-numbers. Where items had to be employed twice, this was done within rather than across modalities.

Visual items were presented for 0·5 sec. (12 frames): the durations of the auditory items averaged 10 frames or 0·42 sec. Materials were presented at three rates: (1) 1 pr./0·5 sec., no interstimulus interval (fast); (2) 1 pr./sec., 0·5 sec. interstimulus interval (medium); (3) 1 pr./2 sec., 1·5 sec. interstimulus interval (slow). There was a 20 sec. interlist interval, with 15 sec. for recall and a ready signal given 0·5 sec. before the start of each list.

Subjects. Forty-five undergraduate psychology students satisfying a course requirement served as subjects. Each subject was assigned to the three shadowing groups (shadow auditory [SA], shadow visual [SV], and no shadowing [NS]) in order of their appearance; there were 15 subjects in each group.

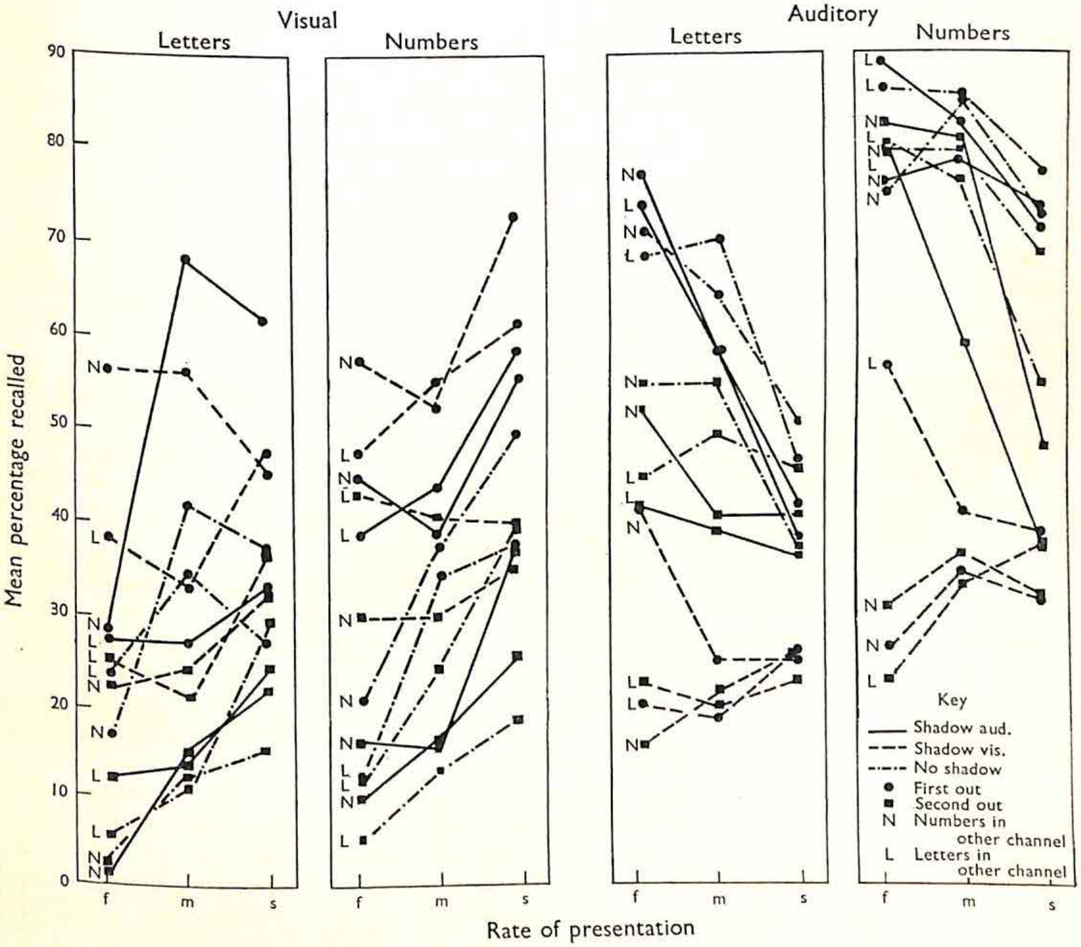
Administration. The experiment was performed in a small room. The projector was in the rear and projected an image on the front wall. All subjects were given general short-term memory instructions, with emphasis placed on the necessity of correct serial order of items in written response. In addition, those subjects in Group SV were told that they were to shadow, i.e. repeat out loud, each visual item as it appeared. Subjects in group SA were instructed to repeat out loud each auditory item as it appeared. The experimenter checked whether the items were repeated correctly. Only one subject in the shadow visual condition shadowed incorrectly and was discarded. Group NS received no additional instructions. The same lists were used for the three groups. The order of presentation of the rate variable was partially counterbalanced, in Latin-square manner. Separate films, each duplicated from the original, were used for each order of rate presentation. On the films the word "auditory" or "visual," indicating the prescribed order of report was spoken immediately after the presentation of the last item in the list; half of the lists were preceded or followed by "auditory" and half by "visual" in a pre-arranged random order.

Each subject was given answer sheets in booklet form—each page represented one rate—with space provided for each response; auditory items were on one half of the page, visual items on the other half. Half of the subjects received answer sheets with visual on the right side, auditory on the left, the other half, the reverse order.

RESULTS

Item analysis. The data for each subject on each string or each sequence of six items were scored according to the accuracy of responses both in nature and location,

FIGURE 1



Intercomparisons of item analysis of bisensory data.

i.e. on an item-by-item basis, permitting individual scores to take on values between 0 to 6. The average scores for each of the 48 tasks are shown in Figure 1. They are expressed as percentages.

The mean performances associated with each stimulus dimension are shown in Table I where the significance tests appropriate to each dimension are indicated.

TABLE I

Means and (Standard Deviations) Associated with Each Dimension of Bisensory Stimulation (Maximum Score = 6)

				d.f.	*test*	*p*
Materials	Numbers: 2·87 (0·57)	Letters: 2·14 (0·50)		44	$t = 12·3$	0·005
Modality	Auditory: 3·09 (1·12)	Visual: 1·91 (0·65)		44	$t = 6·02$	0·005
Order	First: 2·97 (0·70)	Second: 2·65 (0·47)		44	$t = 9·13$	0·005
Shadowing	No Shadowing: 2·72 (0·30)	Shadowing Auditory: 2·65 (0·41)	Shadowing Visual: 2·12 (0·47)	2·42	$F = 9·09$	0·01
Rate	Fast: 2·43 (0·57)	Medium: 2·54 (0·65)	Slow: 2·52 (0·50)	2·88	$F = 1·09$	n.s.

*A fixed effects model is assumed for this analysis and succeeding analyses in which the individual stimulus items are presumed to have equal difficulty. It remains possible, however, that some of the variance associated with experimental treatments ought properly to be assigned to the stimulus strings themselves. In this case all errors would be inflated and the present high significance levels would be somewhat diminished.

These were related sample *t*-tests for the dimensions of modality, order and material; a one-way analysis of variance for shadowing; and a within one-way analysis of variance for rate of presentation (Winer, 1962, p. 112). It is observed that differences in four of the five dimensions had a significant effect on the number of items reported: numbers are retained better than letters; more auditory material is retained than visual material; more items are retained from the channel reported first than from the channel reported second; recall is better for the shadow-auditory and no-shadow groups than it is for the shadow-visual group. The rate variable, however, is not significant. This is understandable from the nature of the significant modality by rate interaction: decreasing auditory recall and increasing visual recall with decreasing rate—thus cancelling the rate effect.

The data were next examined for the purpose of determining interactions among the stimulus dimensions. The possibility was considered of performing a five-dimensional analysis-of-variance but the idea was dismissed. In the first place, it was certain that higher order interactions would emerge from any idiosyncratic features of any of the 48 strings, since each of them represented a unique combination of the materials, rate, modality, and order dimensions.

Therefore each dimension was intercompared with every other dimension by means of two-way analyses of variance. Five interactions proved to be significant: shadowing by rate ($p<0·01$; d.f. = 48, 4): in the NS groups recall was best at the fast rate whereas in the SV group recall was best at the slowest rate; shadowing by modality ($p<0·01$; d.f. = 2, 42): as was expected when the visual channel was shadowed more visual material was retained than auditory, whereas in the NS and SA groups more auditory material was recalled; modality by rate ($p<0·01$; d.f. = 2, 88): auditory strings were retrieved progressively more poorly with decreasing rate of presentation; visual strings, conversely, were best at the slowest rate; modality by material ($p<0·01$; d.f. = 1, 44): although letters were retrieved more poorly than numbers, the disparity was much greater in the auditory channel than it was in the visual channel; modality by order ($p<0·01$; d.f. = 1, 44): while strings recalled first were superior to strings recalled second, the disparity was much greater in the visual channel than it was in the auditory channel.

Further, the data for each modality-materials combination were then examined in terms of the context in which they occurred. Posed as a question, did the nature of the material occurring simultaneously in the other channel affect the success with which items were reported? Since a significant interaction was obtained between modality and material, the question was examined separately for each modality-material combination. The mean recall scores for like and unlike accompaniments are presented in Table II. For three of the combinations, related-sample *t*-tests indicate that performance is independent of the nature of the material in the other channel. For visual letters, however, a significant disparity obtains. Thus, the retrieval of visual letters appears to be suppressed by the concurrent presentation of letters in the auditory channel.

TABLE II

MEAN CONTEXT RESPONSES AND (STANDARD DEVIATIONS) FOR THE ITEM ANALYSIS (d.f. = 44)

Modality-Material	*Accompaniment*		t	p
	Like	*Unlike*		
Numbers-Auditory	3·54 (1·41)	3·71 (1·37)	<1	——
Letters-Auditory	2·46 (1·12)	2·64 (1·07)	<1	——
Numbers-Visual..	2·22 (0·76)	1·99 (0·91)	<1	——
Letters-Visual	1·51 (0·67)	1·92 (0·76)	1·3	0·05

DISCUSSION

General effects. Some of the variables studied produced effects in bisensory memory similar to those recorded for them in the existing literature on unisensory memory and other investigations of bisensory memory and many of the effects obtained are self-explanatory.

For example, it was not unexpected that the channel reported first should be the channel recalled best simply because it is first out, and therefore not subject to lengthy storage demands.

There were also material differences which were not without a precedent. Thus, while the letters variable was defined by 10 items these 10 items were part of a potentially larger response set consisting of all 26 letters of the alphabet. There is no assurance that instructions restricting the stimulus set served to adequately restrict subjects' response set (Garner, 1962). By contrast, the 10 digits defined both the stimulus and response set. Therefore, it is not surprising that numbers prove considerably easier to recall than did letters under any condition of short-term memory.

On the other hand, there were variables that yielded results which were unexpected on the basis of the existing literature. For example, it was demonstrated that the material in one channel generally did not influence recall in the other channel; material combinations were irrelevant in determining recall. This, of course, is not in keeping with retroactive and proactive interference data.

Modality, rate, and shadowing. The most relevant results in terms of the immediate interests of this study were the modality, rate, and shadowing effects to be discussed in turn.

The modality effect cannot be considered independently of rate of presentation. Bisensorily, both modalitites displayed curves predictable from data obtained for each modality when it was explored independently of the other modality. Auditory results yielded a decay function with decreasing rate; visual results yielded an increasing function. Thus the integrity of the modality-rate was maintained even when both modalities were stimulated simultaneously.

It was previously found (Dornbush, 1966, 1967) bisensorily and unisensorily that the visual function increased with respect to decreasing rate of presentation and it was suggested that this was so possibly because of the temporal delay involved in transforming visual items into auditory subvocalizations. That is, with increasing intervals between successive items, there is increased opportunity to translate and rehearse visual materials. If this were the case, then there might be one type of store—an auditory store. If in fact this translation or transformation were impeding recall, then providing as "simultaneous" or "immediate" vocalization or transformation as possible might alter the type of visual function obtained so that it perhaps approximated the auditory function. This thought formed the basis of this experiment.

However, the results of the present experiment indicate that immediate vocalization via shadowing of a modality did not alter the visual function in the bisensory situation. Indeed, recall still increased with a decrease in presentation rate. Thus while visual input no doubt enters an auditory store (Conrad, 1962, 1964; Sperling, 1963) it is not untenable to assume that it is a store somewhat different from the store to which items that are initially auditory terminate. Such a suggestion is in line with Margrain's (1967) statement that short-term storage is modality-specific. Specifically, she suggests "that there may initially be two short term stores: one for visual material and one for auditory, since interpolated activity of one kind damaged the retention of earlier material of that kind, more than material of the alternative type" (p. 113). She further states that while Broadbent's S and P mechanisms are not descriptive of processing of simultaneously presented material even Sperling's suggestion of a final auditory store in both auditory and visual material is unsupported.

One of the salient findings resulting from the shadowing procedure is that equal division of attention between the auditory and visual modality is in fact not possible. Figure 1 demonstrates that there is very little difference in recall in the no-shadow and shadow-auditory conditions for both modalities. Obviously, then, near maximum attention must be given the auditory modality under the no-shadow or ordinary bisensory condition and at the same time minimum attention given to the visual modality under the no-shadow condition as recall in this modality is not less under the shadow-auditory condition.

Thus it must be concluded that in the ordinary bisensory simultaneous condition attention to the visual modality is near minimum while auditory attention is maximal. That is, it is thought that the amount recalled visually is almost incidental to the task.

The increase in visual retention in the SV group can be due either to (1) shadowing forcing attention to materials in the visual channel or (2) to the conversion of visual signals to the mode in which they are stored or of course to a combination of the two. The indications are, at present, that increased visual retention with bisensory presentation when that modality is shadowed is attributable to increased *attention* to that modality. Previous experimentation (Dornbush, unpublished data) with shadowing of visual material when it is presented independently of auditory material indicates no particular advantage for recall over the non-shadowed condition.

This work was supported by a research grant from the United States Public Health Service, (NIH MH 12657). Portions of this paper were read at the 75th Annual Convention of the American Psychological Association, Washington, D.C., September, 1967.

REFERENCES

BROADBENT, D. E. (1963). Flow of information with the organism. *J. verb. Learn. verb. Behav*, **2**, 34–9.

CONRAD, R. (1962). An association between memory errors due to acoustic masking of of speech. *Nature*, **193**, 1314–5.

CONRAD, R. (1964). Acoustic confusions in immediate memory. *Brit. J. Psychol.*, **55**, 75–84.

DORNBUSH, R. L. (1966). Input variables in bisensory memory. *Proceedings of the XVIIIth International Congress of Psychology*. Moscow, USSR.

DORNBUSH, R. L. (1967). *Input Variables in Bisensory Memory*. Unpublished doctoral dissertation. City University of New York.

GARNER, W. R. (1962). *Uncertainty and Structure as Psychological Concepts*. New York: Wiley.

MARGRAIN, S. A. (1967). Short-term memory as a function of input modality. *Quart. J. exp. Psychol.*, **19**, 109–14.

SPERLING, G. (1960). The information available in brief visual presentations. *Psychol. Monogr.*, **74**, No. 11.

SPERLING, G. (1963). A model for visual memory tasks. *Human Factors*, **5**, 19–31.

WICKELGREN, W. A. (1965). Acoustic similarity and intrusion errors in short-term memory. *J. exp. Psychol.*, **70**, 102–8.

WINER, B. J. (1962). *Statistical Principles in Experimental Design*. New York: McGraw-Hill.

Manuscript received 15th December, 1967.

REPETITION EFFECTS AND SIGNAL CLASSIFICATION STRATEGIES IN SERIAL CHOICE-RESPONSE TASKS

BY

PATRICK M. A. RABBITT*

From the M.R.C., Applied Psychology Research Unit, 15 Chaucer Road, Cambridge

Repetition-effects (Bertelson, 1965) were examined in three serial self-paced choice-response tasks in which each response was made to all members of a class of more than one signal, and in one task in which eight different responses were each made to one of eight different signals. Three kinds of transition between successive responses occur in such tasks: transitions between *Identical responses* where the same signal and response are immediately repeated, transitions between *Equivalent responses* where the same response is made to a new signal and transitions between *New responses* where neither signal nor response are repeated. The relative reaction-times for these three classes of events were found to vary as a function of stimulus information load, as a function of response information load and as a function of the level of practice which subjects attained in the task in question. These variations allow some comment on the utility of recent models for serial and parallel stimulus analysis as explanatory constructs for the repetition-effect.

INTRODUCTION

In serial choice-response tasks human subjects respond faster if a given signal, and the response to it, are immediately repeated than if they have to identify a new signal and to program a new response (Hyman, 1953). This has become known as the "repetition effect" (Bertelson, 1961; 1965).

A similar, facilitating effect is observed when subjects make successive, different movements with the same limb to successive, different signals (Rabbitt, 1965). It is therefore likely that two sets of factors contribute to the "repetition effect" observed by Bertelson and Hyman; that is factors which relate to the repetition of a particular signal and other factors which relate to the repetition of a particular motor act. It becomes a real question how the proportionate contributions of these factors to the overall repetition-effect can be estimated.

A test case is a task in which the subject makes the same response to all members of one set of signals and makes other responses to signals in other arbitrary sets. In such a task the number of signals between which the subject has to discriminate, and the number of responses between which he has to choose, may be varied independently (Rabbitt, 1959). In most previous studies (Hyman, 1953; Bertelson, 1961; Falmagne, 1965; Leonard, Newman and Carpenter, 1966), the effects of signal and response entropy upon repetition-effects have been confounded because the repetition of a signal has always entailed the repetition of a particular motor act. An exception is an experiment by Bertelson (1965) who compared RTs for the three classes of transitions between successive signals and responses which are possible in a task in which subjects have to choose between two responses, each of which is appropriate to two of four possible signals (2R/4S mapping) which occur during a serial self-paced task. For convenience these classes of transitions may be defined and labelled as follows:

Identical transitions when the same signal, and implicitly the same response, are immediately repeated. *Equivalent transitions* when a signal is followed by a different signal to which the same response is appropriate and *New transitions* when a signal is followed by a different signal to which a new response must be made. Bertelson found that four practised subjects gave longer RTs for New transitions than for Equivalent or Identical transitions. For two of these subjects RTs for Equivalent transitions were slightly longer than for Identical transitions, while for the remaining two subjects Equivalent transition RTs were equal to Identical transition RTs.

---

* Now at the Institute of Experimental Psychology, 1 South Parks Road, Oxford.

Unfortunately it is not possible to derive a general model for signal-and-response repetition effects from Bertelson's results since he sampled a single level of signal and response entropy at a single point in practice. Previous work had suggested that the effect of signal and response entropy interact multiplicatively to determine choice RT (Rabbitt, 1959; Pollack, 1963) and that the nature of this interaction may change with practice (Rabbitt, 1962). It seemed that an investigation of variations in the proportional contributions of H_s and H_r to the total repetition effect at different points in practice might provide a useful tool for testing current models which undertake to describe the process of discrimination between classes of signals and the selection of responses to them. An experiment was therefore made to investigate the relative changes in RTs for Identical, Equivalent and New transitions, with practice, in four tasks between which H_s and H_r were systematically and independently varied.

EXPERIMENT

Subjects. Forty-five Royal Navy ratings aged from 18 to 27 served as subjects.

Apparatus and procedure. The experiment was made with a stimulus presentation and reaction-timing apparatus (SPARTA). This equipment was programmed with punched tape to present sequences of signals (digits) on a "Digitron" G.S.R. 10 J numerical display tube. The subject answered each signal in turn by pressing one of eight keys which were inset into a desk before him so as to rest conveniently under the four fingers of each hand. Within 20 msec. of the closure of a microswitch under any of these keys SPARTA presented a new signal on the Digitron tube. SPARTA then punched out on 5-channel tape the codes which specified the signal presented to the subject, the key he pressed in answer to it and the elapsed time since the last key closure (to within 10 msec.).

In all conditions an experimental run with SPARTA was a sequence of 301 signals and responses in a self-paced mode. In all experimental conditions all subjects experienced two such runs on each of five successive weekday mornings.

The same ten programmes of signals were used for three conditions of the experiment. These consisted of the digits 1 to 8 programmed from tables of random numbers with constraints to ensure that, as far as possible, all 64 transitions between signals occurred equally often. A different one of these sequences was used for each of the ten runs experienced by each subject. Three separate groups of subjects mapped responses on to these eight signals in the following ways:

2R/8S condition: 12 subjects responded to any of the digits 1, 2, 3, or 4 by pressing a key inset under their left forefinger, and to any of the digits 5, 6, 7 and 8 by pressing a key inset under their right forefinger.

4R/8S condition: 11 subjects responded to the digits 1 and 2 with their left middle fingers, to the digits 3 and 4 with their left fore-fingers, to the digits 5 and 6 with their right fore-fingers and to the digits 7 and 8 with their right middle fingers.

8R/8S condition: 11 subjects responded to the eight decimal digits using one of their eight fingers for each digit.

The allocation of digits to fingers was in numerical order reading across from left little finger (1) to right little finger (8).

Ten similar programmes of 301 signals were made up so that the digits 1 to 4 occurred in random order, with approximately equal numbers of transitions between signals. These were used to provide a further condition:

2R/4S condition in which subjects responded to digits 1 and 2 with their left fore-fingers and to digits 3 and 4 with their right fore-fingers.

In all these conditions the apparatus never gave the subject any indication whether any of his responses was correct or wrong. The recording of signals and responses nevertheless allowed the experimenter to detect errors when they were made.

RESULTS

Errors

A teleprinted record was made of each subject's output tapes for the second and tenth practice runs in each condition. These records were analysed by eye to locate errors. The percentages of errors made at each level of practice sampled in each condition are set out in Table I below.

TABLE I

MEAN PERCENTAGE OF INCORRECT RESPONSES MADE BY SUBJECTS EARLY AND LATE IN PRACTICE ON EACH OF FOUR SERIAL, SELF-PACED CHOICE-RESPONSE TASKS

Group		*Responses* 301 *to* 600	*Responses* 1,201 *to* 1,500
2R/4S	Two responses, four signals	7·2	3·4
2R/8S	Two responses, eight signals	8·4	2·9
4R/8S	Four responses, eight signals	6·9	3·6
8R/8S	Eight responses, eight signals	8·8	3·2

A separate analysis of variance was made at each level of practice on the error-scores over all conditions. Differences between conditions were significant on the second practice run ($p < 0{\cdot}01$) but not on the tenth practice run ($p < 0{\cdot}1$).

Reaction times

Errors, and the three responses following each error, were known to be atypical (Rabbitt, 1966) and so were left out of the analyses described below. The remaining correct responses were divided into three classes: *Identical transitions* where the identical signal and the response to it were repeated, *Equivalent transitions* where a signal was followed by another in the same response class (so that the response was repeated though the signal was not), and *New transitions* where neither the signal nor the response was repeated.

Each subject's mean RT for each of these classes of response was calculated from the SPARTA print-out. Means of these means, for each level of practice, are set out in Table II. To render comparisons between conditions more accurate RTs for Identical responses made with the fore-fingers of each hand are set out for comparison between conditions.

Two analyses of the data were undertaken at each level of practice: RTs were first compared between transition classes within each condition and then RTs within each transition class were compared between conditions. Data for the second and tenth practice runs are discussed separately.

Second practice run. t-tests were used to analyse the data within each condition. In all conditions Identical transition RT < New transition RT ($p < 0{\cdot}01$). Within the 2R/4S, 2R/8S and 4R/8S conditions Identical RT < Equivalent RT ($p < 0{\cdot}01$ in all cases). However, in marked contrast to previous findings (Bertelson, 1965) Equivalent transition RT was not significantly different from New transition RT in any condition ($p > 0{\cdot}1$).

In order to analyse differences between conditions within each transition class an Analysis of Variance was made, and the residual term was used to calculate S^2. New transition RTs were significantly different across conditions (An.o.Va. $p < 0{\cdot}001$). S^2 showed that 8R/8S > 4R/8S > 2R/8S > 2R/4S ($p < 0{\cdot}01$ in each case).

TABLE II

MEAN RTs FOR THREE CLASSES OF TRANSITIONS BETWEEN SIGNALS AND RESPONSES. DATA FROM FOUR SERIAL, SELF-PACED CHOICE-RESPONSE TASKS AT TWO-LEVELS OF PRACTICE

		Responses 301 to 600		*Responses 1,201 to 1,500*		*Fore-finger RTs only. Responses 1,201 to 1,500*	
Group		*Mean*	*s.d.*	*Mean*	*s.d.*	*Mean*	*s.d.*
2R/4S ($n = 11$)	New signal/New response	584	219	424	182		
	Equivalent signal/Same response	567	231	394	126		
	Identical signal/Same response	491	184	383	142		
2R/8S ($n = 12$)	New signal/New response	608	251	484	197		
	Equivalent signal/Same response	580	237	416	159		
	Identical signal/Same response	510	219	377	162		
4R/8S ($n = 11$)	New signal/New response	725	270	592	201		
	Equivalent signal/Same response	720	284	532	215		
	Identical signal/Same response	652	247	416	224	386	183
8R/8S ($n = 11$)	New signal/New response	970	291	770	234		
	Identical signal/Same response	693	252	461	209	429	214

Equivalent transition RTs were also significantly different between conditions (An.o.Va. $p < 0{\cdot}001$). Again S^2 gave 4R/8S > 2R/8S > 2R/4S ($p < 0{\cdot}01$ in each case).

Identical transitions RTs were significantly different between conditions (An.o.Va. $p < 0{\cdot}01$). S^2 gave 8R/8S > 4R/8S > 2R/8S ($p < 0{\cdot}01$). The difference between 2R/8S and 2R/4S was not significant.

Tenth practice run. *t*-tests were again used to analyse differences between transition-classes within each condition. In all conditions New RT > Identical RT ($p < 0{\cdot}01$) and (where applicable) New RT > Equivalent RT ($p < 0{\cdot}01$). In contrast to the data for the second practice run (where the difference between New RT and Equivalent RT was not significant) on the final run in the 2R/4S, 2R/8S and the 4R/8S conditions New RT > Equivalent RT ($p < 0{\cdot}01$). In the 2R/4S and 4R/8S conditions Equivalent RT > Identical RT ($p < 0{\cdot}05$).

Analysis of Variance and S^2 were again used to analyse differences between conditions within each transition class. New RT was significantly different between conditions ($p < 0{\cdot}001$ and An.o.Va.). S^2 gave 8R/8S > 4R/8S > 2R/8S or 2R/4S (at $p < 0{\cdot}01$) and New RT 2R/8S > New RT 2R/4S (at $p < 0{\cdot}05$). The same trend appeared in Equivalent RTs (An.o.Va. $p < 0{\cdot}01$). That is, 4R/8S > 2R/8S ($p < 0{\cdot}01$) and 2R/8S > 2R/4S ($p < 0{\cdot}05$).

In contrast to New transition RTs and Equivalent transition RTs, Identical transition RTs did *not* differ significantly between conditions (An.o.Va. $p > 0{\cdot}1$). However, on using Mann-Whitney U tests to compare Identical transition RTs between conditions, it emerged that Identical transition RTs in the 8R/8S condition were significantly longer than in any other ($p < 0{\cdot}05$ to $p < 0{\cdot}01$ two-tailed). No other differences between conditions were found ($p > 0{\cdot}2$).

Discussion

The results of these experiments show that the proportions of the overall repetition effect attributable to the repetition of signals and to the repetition of responses do not simply vary with H_s or with H_r, but rather with some complex interaction between these two factors. The nature of this interaction changes sharply with practice.

It is fashionable to describe the classification of complex signals in terms of models derived from computer programs for character-recognition. Let us consider the present data in terms of two classes of these models respectively put forward by Sternberg (1966) and by Neisser (1963; 1964). The main features of the present results, for which any model must account, may be listed as follows:

Identical transition RTs vary with both H_s and H_r on the second practice run. But by the tenth practice run variations in H_s and H_r affect Identical RT very little (8R/8S) or not at all (2R/4S; 2R/8S; 4R/8S). Any model must explain this change in practice, and also explain why Identical transition RT provides an exception to the general rule that $RT = f(H_s \text{ and } H_r)$.

Equivalent transition RTs vary sharply with the number of response sets into which the signal ensemble is partitioned (2R/8S vs. 4R/8S) and to a lesser extent with the number of signals in each response set (2R/4S vs. 2R/8S). This is true at both levels of practice examined.

The magnitude of differences between New, Equivalent and Identical transition RTs varies with the level of practice. On the second practice run Equivalent transition RTs are not significantly different from New transition RTs. However, Equivalent transition RTs reduce more sharply with practice than do either New or Identical transition RTs, so that by the tenth practice run Equivalent transition RTs are within 11 msec. (2R/4S) to 30 msec. (2R/8S) of Identical transition RTs. Because the New, Equivalent and Identical transition RTs are differentially altered by practice it is obvious that no single "steady state" model can adequately fit all the data. Let us therefore consider results for the second and for the tenth practice runs separately:

During the tenth practice run, on Identical transitions, the subjects seem to be able to respond without considering any of the other possible signals which they may have to identify (viz. 2R/4S vs. 2R/8S) or any of the other possible responses between which they may have to choose (viz. 2R/8S vs. 4R/8S). To account for repetition effects Bertelson (1963) suggested that subjects carrying out serial choice-response tasks always begin their analysis of each successive signal by making a test to determine whether or not it is the same as the last signal to which they responded. Bertelson suggested that if such a test is positive the subject can immediately repeat his last response without going on to test for any of the remaining signals in the ensemble. Thus, whatever the size of the signal and response ensembles, Identical RT should always be the time required to make a single test on perceptual input, and to repeat the same motor act.

Recent descriptions of perceptual analysis in terms of computer routines for character identification (Neisser, 1963; Sternberg, 1966) allow us to re-state Bertelson's suggestion as one of several models which we may test against the present results.

Computer routines which identify complex signals may be broadly sub-divided into serial testing routines, parallel testing routines or a combination of the two (hybrid routines).

A *parallel routine* is conceived as simultaneously making a set of X_t independent tests on input, each of these tests establishing whether the input can, or cannot, be classified as one of X_I different states (e.g. Neisser, 1963; 1964).

A *serial routine* would also solve the same problem by making X_t independent tests, but would make them one at a time in succession.

A *hybrid routine* would also make X_t tests, but would do so by successively making batches of parallel tests for *n* different sub-groups of states.

Serial and hybrid routines may be further broken down, since they may be either self-terminating or exhaustive.

An *exhaustive routine* always continues until all possible X_t tests have been made before input is classified.

A *self-terminating routine* only continues testing until any test, or set of tests, is successful. At this point input is classified without any further tests being made. Bertelson's (1965) suggestion that the subject always repeats a test for the last signal which he identified can therefore be re-stated as an hypothesis that the subject's perceptual analysis is a serial or hybrid self-terminating process—with the additional assumption that the first test is made for the signal identified on the last trial. Let us see how this model fits the data for each level of practice.

Data from the tenth practice trial allow us to discount models based on either parallel or serial exhaustive routines. If the subject always tests for all signals in parallel there is no reason why repetition of an identical signal should give faster RTs than repetition of an Equivalent signal (the same motor response is repeated in either case). For the same reason we can reject serial exhaustive models, which also suppose that, before a decision is made, tests must be made for *all* signals in turn (whether in the entire ensemble, or within a given sub-set of signals within that ensemble). It seems that the subjects use a serial self-terminating routine—but the present data do not allow us to say whether this is an hybrid routine or not. With an hybrid routine Identical RTs would be faster than Equivalent RTs, if the subject repeated, to each signal, a set of X_t tests; but where X_t did not coincide with any set of signals which included the test which had identified the preceding item. An exception to this would occur when the sub-set of tests X_t coincided exactly with the limits of a set of "Equivalent" signals (as defined for the subject by the experimenter). In this case Equivalent and Identical transition RTs would be the same, since repetition of any test would necessarily imply repetition of tests for all Equivalent signals. Thus, at least on the tenth practice run, we may assume that if subjects *do* use a hybrid testing routine, the sub-sets of tests which they make do not exactly coincide with those into which the experimenter partitioned the signal set.

In the second practice run, in all conditions, Identical transition RTs were again faster than Equivalent transition RTs. As we have seen, we can only account for this by assuming that they use a serial or hybrid self-terminating routine. It is tempting to conclude that subjects use the same broad class of testing routine both early and late in practice—but this statement would leave the most striking features of the data unexplained. Early in practice Identical transition RT varies with the number of signal and response alternatives among which the subject has to select, and Equivalent transition RTs are not significantly different from New transition RTs. We must conclude that when an Identical transition occurs the subject must, at least on some occasions, consider other signal and response alternatives before making a decision. At this stage in practice, in marked contrast to Bertelson's (1965) findings (and to the data from the tenth practice session) the repetition of a response seems to facilitate RT much less than the repetition of a *signal*. It seems that practice does not merely shift the overall distribution of RTs, but selectively affects the rank order of transition-classes within this overall distribution. Any model we propose must account for this fact. We may choose between different models, depending on whether we assume that subjects use an hybrid routine or a serial self-terminating routine.

A serial self-terminating routine assumes that the first test of a series is always for the signal which had last been successfully identified—but implies no assumption about the order in which further tests are made. It is possible to suppose that, with practice, subjects progressively adapt the order of testing to coincide with the experimenter's partitioning of the signal ensemble. That is, having unsuccessfully tested for the last signal which he identified, we may suppose that the subject systematically tests through all signals in the Equivalent response class before considering (New) signals in other classes. A gradual trend in this direction during practice would progressively alter the distributions of Equivalent transition RTs so that they increasingly approximated to Identical transition RTs (while never, on this hypothesis, becoming as fast as Identical transition RTs).

If we assume that subjects use an hybrid self-terminating testing routine it also does not follow that the sub-sets of parallel tests which we suppose the subject makes (and which we suppose that he repeats when any one of them successfully identifies a signal) coincide with the sub-sets into which the experimenter partitions the signal ensemble. It is possible that one effect of practice is to progressively modify the sub-sets of parallel tests made by subjects until they coincide with the experimenter's partitioning of the signal ensemble. This would progressively alter the distributions of RTs within transition-classes, so that Equivalent transition RTs gradually approximate to Identical transition RTs (and become equal to Identical transition RTs at some limiting level of practice).

Both these hypotheses beg the question as to *why* the subject should modify his testing strategy so as to be able to respond relatively quickly when Equivalent transitions occur. A possible answer lies in the fact that the magnitude of the repetition effect reduces sharply as the interval between successive responses increases (Bertelson, 1961; 1965). Any reduction in the time taken for the perceptual identification of Equivalent signals may therefore allow the subjects to benefit from facilitation of RT due to repetition of motor components of a response. On this line of argument the absence of a *response* repetition effect on Equivalent transitions early in practice might mean that identification and classification of signals takes so long that successive responses are too widely separated in time for maximum facilitation to occur.

Either a serial self-terminating or an hybrid testing model can also serve to explain another, apparently anomalous, feature of these results. In the 2R/4S condition, where each response class has two members, the difference between Identical and Equivalent RT is 11 msec. We must therefore explain why this difference is so much larger (126 msec.) in the 4R/8S condition, where response classes are of the same size. Whether we base explanations for practice effects on a serial or on an hybrid self-terminating model, we assume that subjects gradually adapt their testing routines to the experimenter's partitioning of the signal ensemble. Previous investigations of RTs to classes of signals have shown that the time taken to learn a partition of a given signal ensemble is proportional to the number of sets into which it is classified (Rabbitt, 1959; 1962; Pollack, 1963). It would be consistent with these findings that a 4R/8S classification should take longer to learn than a 2R/8S classification—and that the approximation of Equivalent transition RTs to Identical transition RTs is delayed in the 4R/8S condition for this reason.

Previous studies of signal classification give results of two kinds. On the one hand experiments with relatively unpractised subjects have shown that RT = f(H_s) when H_r is constant (Nickerson and Feehrer, 1964). This may amount to a linear relationship between RT and relevant set size when subjects briefly experience many different partitionings of a single ensemble of signals during a relatively short experiment (Sternberg, 1966). On the other hand, data from highly practised subjects suggest that the effects of H_s upon RT diminish with practice (Rabbitt, 1962) until there may be little (Rabbitt, 1959; Pollack, 1963) or no (Broadbent and Gregory, 1962) measurable effect of signal entropy (H_s) on choice RT. In visual search, where subjects discriminate between two sets of (relevant and irrelevant) signals, it has been specifically claimed that, at some limiting level of practice, search time becomes independent of relevant set size because subjects gradually learn to adopt a parallel routine, and to test simultaneously for all relevant items (Neisser, 1963; 1964).

The present results support the view that practice does not merely bring about a gross reduction in RT, but rather alters the distribution of RTs in a way which can only be interpreted on the assumption that strategies of perceptual analysis are gradually modified as subjects learn a given classification. The present results preclude the possibility that parallel processing occurs by the tenth practice session. Neither these results (—nor perhaps any others!) can preclude the possibility that at some undefined limiting level of practice subjects may be able to change from a serial to a parallel hybrid testing strategy. Indeed, some support for this view is the fact that three out of 12 subjects in the 2R/4S condition and two out of the 12 subjects in the 2R/8S condition showed no significant difference between Identical transition RT and Equivalent transition RT on their tenth practice runs. The same was also true for two out of four subjects highly practised by Bertelson (1965) in a 2R/4S task.

The most general conclusions of this paper thus reinforce those of a previous study of the ways in which practice allows subjects to optimize their techniques of discrimination between sets of complex stimuli (Rabbitt, 1967). Unlike poorly programmed computing systems of limited capacity, human beings can apparently use many alternative routines to classify perceptual input. If we attempt to describe performance in terms of some unique classification routine (Sternberg, 1966; Neisser, 1963; 1964) it is very possible that our model can be validated—but only at a carefully selected level of practice. A more realistic, and interesting, research strategy is surely the investigation of the subject's progression through a series of classification routines until he evolves one which reflects the optimal relation of his information-processing capacity to the task which we require him to perform.

REFERENCES

BERTELSON, P. (1961). Sequential redundancy and speed in a serial two-choice responding task. *Quart. J. exp. Psychol.*, **13**, 90–102.

BERTELSON, P. (1963). S–R relationships and reaction times to new vs. repeated signals in a serial task. *J. exp. Psychol.*, **65**, 478–84.

BERTELSON, P. (1965). Serial choice reaction-time as a function of response versus signal-and-response repetition. *Nature*, **206**, 217–8.

BROADBENT, D. E., and GREGORY, M. (1962). Human response to classes of stimuli. *Nature*, **192**, 1315.

FALMAGNE, J. C. (1965). A linear model for choice R.T. with application to experimental results. *J. math. Psychol.*, **2**, 77–127.

HYMAN, R. (1953). Stimulus information as a determinant of reaction-time. *J. exp. Psychol.*, **45**, 188–97.

LEONARD, J. A., NEWMAN, R. C., and CARPENTER, A. (1966). On the handling of heavy bias in a self-paced task. *Quart. J. exp. Psychol.*, **18**, 130–141.

NEISSER, U. (1963). Decision-times without reaction-time: Experiments in visual scanning. *Amer. J. Psychol.*, **76**, 376–85.

NEISSER, U. (1964). Visual search. *Scient. Amer.*, **210**, 94–102.

NICKERSON, R. S., and FEEHRER, C. E. (1964). Stimulus categorisation and response-time. *Percep. mot. Skills.*, **18**, 785–93.

POLLACK, I. (1963). Speed or classification of works into superordinate categories. *J. verb. Learn. verb. Behav.*, **2**, 159–65.

RABBITT, P. M. A. (1959). Effects of independent variations in stimulus and response probability. *Nature*, **183**, 1212.

RABBITT, P. M. A. (1962). *Perceptual Discrimination and the Choice of Responses.* Unpub. Ph.D. thesis. Univ. of Cambridge.

RABBITT, P. M. A. (1965). Response-facilitation on repetition of a limb-movement. *Brit. J. Psychol.*, **56**, 303–4.

RABBITT, P. M. A. (1966). Errors and error correction in choice response tasks. *J. exp. Psychol.*, **71**, 264–72.

RABBITT, P. M. A. (1967). Learning to ignore irrelevant information. *Amer. J. Psychol.*, **80**, 1–13.

STERNBERG, S. (1966). High-speed scanning in human memory. *Science*, **3736**, 652–4.

Manuscript received 2nd February, 1968.

CHANNEL-CAPACITY, INTELLIGIBILITY AND IMMEDIATE MEMORY

BY

PATRICK M. A. RABBITT*

From the Medical Research Council Applied Psychology Research Unit, Cambridge

When subjects try to remember lists of digits played to them through pulse-modulated white noise the number of errors they make is greater than would be expected if digit-recognition errors and immediate memory errors were independent (Exp. 1).

A second experiment compared recall of digits in early list positions, when digits in subsequent list positions were presented through noise, and in clear. Digits in early positions were less well remembered when digits in later list positions had to be discriminated through noise.

In a third experiment prose passages were played to subjects who subsequently answered questions about their factual content. Judged by this technique recall of the first half of a prose passage is less accurate if the second half must be heard through noise than if the entire passage is heard through a good fidelity system. These results together are interpreted as demonstrations that increased difficulty of recognition of speech through noise may interfere with other activities, (conveniently termed "rehearsal") which may be necessary to efficiently retain data in memory.

INTRODUCTION

Words transmitted through communications links, such as telephones, may be distorted by the characteristics of the system or partially masked by noise. It has been shown that there are at least two different ways in which noise or distortion can reduce the efficiency with which speakers can communicate with each other.

(1) The noise may make it impossible, or very difficult, to recognize all the words presented. This fact is obvious, and extremely well attested (Kryter, 1960). It is now possible to specify which bands of the frequency-intensity spectra convey the different kinds of information necessary in various semantic and syntactic contexts—and consequently to judge the extent to which the perception of the cues which are used to recognise words are modified by various kinds of masking or distortion (Webster and Klumpp, 1965).

(2) When masking or distortion is not sufficient to produce errors of recognition a listener may nevertheless have to make more effort to distinguish what is said to him. This may imply that a listener who can correctly report words presented over a degraded circuit may be less competent if he is required to simultaneously perform other unrelated tasks (Broadbent, 1958). In an information-theory framework we may suppose that in order to recognize degraded speech listeners have to use "spare channel capacity" which they could otherwise distribute between two tasks so as to maintain efficiency on both (Poulton, 1958; Brown, 1964).

Many experiments have shown that two unrelated tasks may compete to pre-empt a single channel of limited capacity (Poulton, 1958; Brown, 1964). It is less clear how such a model would operate in a case where one of two simultaneous tasks requires the subject to recognize words presented to him through noise, while the "secondary task" consists in remembering the words which are recognized, or in performing some conceptual transformations upon them. Such a task emphasises the logical difficulties in defining what a subject has achieved when he is said to have "recognized" a sequence of words. Does he recognize the sense of the words in that he can supply synonyms for them? Can he remember the words in the order in which they were presented? Or can he simply mimic the sounds which he heard? If we assume that some, or all, of these may be empirically separable processes (Chistovitch, 1965) we can also enquire whether they compete for channel capacity in the same way as two or more simultaneously conducted perceptual-motor tasks.

---

*Now at Institute of Experimental Psychology, 1 South Parks Road, Oxford.

Recent experiments (Posner and Rossman, 1965) suggest that items are maintained in immediate memory by some activity, conveniently termed "rehearsal," which requires part of the subject's channel-capacity. Any activity intervening between the presentation of items to the subject and his subsequent reproduction of them is assumed to occupy channel capacity, inhibiting rehearsal and reducing the probability that items will be correctly recalled. Posner and Rossman (1965) showed that interpolated secondary tasks of high information-load produced greater decrement in recall than low information-load tasks. Variations in the information-load of signals may be expressed in terms of their mutual discriminability (Crossman, 1955; Laming, 1968). The question therefore arises whether recognizing some words may not be regarded as a secondary task pre-empting channel-capacity required for the retention of other words—with the corollary that when recognition is made more difficult by noise, interference with rehearsal may be correspondingly greater.

Two experiments were made to determine whether items which are difficult to recognize are also less easy to remember. Experiment 1 is a direct test for an interaction between reduced discriminability of presented items and the probability that they will be correctly recalled. It was made to discover whether levels of noise which have little or no effect on performance when lists of digits have to be recognized and transcribed nevertheless significantly reduce the efficiency with which they can be stored and recalled.

EXPERIMENT 1

Method

Forty lists of eight digits in random order without repeats were recorded by an experienced male speaker at the rate of 1 digit/sec. with 10 sec. pauses between successive lists. In the experimental design these were considered as four successive groups of ten lists (A_1, A_2, B_1 and B_2). The original tape was re-recorded through a simulated G.P.O. telephone-link incorporating a Modulated Noise Reference Unit (M.N.R.U.). This latter device enables noise in the band 0 to 3·5 kc/s. to be added to the speech signal in such a manner that the intensity of the added noise is directly proportional to the instantaneous speech amplitudes. The noise intensity therefore varies with the speech-signal, maintaining a constant signal-to-noise ratio.

The master list was re-recorded through this link four times to give the following paradigms where (N) indicates that digits were recorded with the M.N.R.U. set to give a OdB. signal to noise ratio and (C) indicates that the M.N.R.U. was switched out of the link and that the list was recorded with the maximum fidelity which the system allowed.

(1) A_1(N) A_2(C) B_1(N) B_2(C);
(2) A_1(N) A_2(C) B_1(C) B_2(N);
(3) A_1(C) A_2(N) B_1(N) B_2(C);
(4) A_1(C) A_2(N) B_1(C) B_2(N);

Each of these tapes was played to two separate groups of 10 subjects. One of these groups transcribed the A lists and remembered and reproduced the B lists while the other reversed the order of conditions. The tape-recorder was stopped after every 10 lists, when the subjects were told whether the next 10 lists were to be read through noise or in clear.

Subjects

Subjects were 36 men and 44 women volunteers from the A.P.R.U. civilian panel. They were aged from 26 to 68 years (Mean = 47.2 years).

Results

Lists were scored correct only if all digits presented were reported in the correct order. The mean number of lists correctly reproduced in each Condition is set out in Table I.

TABLE I

MEAN NUMBER OF LISTS OF EIGHT DIGITS CORRECTLY REPRODUCED
EXPERIMENT 1.

	Digits presented in clear	*Digits presented in with OdB. signal to noise ratio*
Transcription	10·00 ($\sigma = 0$)	9·64 ($\sigma = 0{\cdot}48$)
Memory	4·02 ($\sigma = 3{\cdot}9$)	2·84 ($\sigma = 4{\cdot}2$)

Analysis of variance gave significant terms for Order ($p<0{\cdot}01$) for differences between subjects ($p<0{\cdot}01$) Transcription vs. Memory ($p<0{\cdot}001$) and for Noise ($p<0{\cdot}001$). The interaction term for Noise vs. Memory was significant ($p<0{\cdot}01$).

Discussion

Lists of eight digits are less likely to be correctly recalled when they are presented through noise than when they are presented in clear. The significant Memory/Noise interaction discounts the simple hypothesis that the recall-errors for degraded digits represent the simple sum of the (estimated) recognition errors and of the recall errors when digits are presented in clear.

Two different explanations are possible. Lists of items are less well-remembered if they are acoustically confusable with each other than if they are acoustically distinct (Conrad, 1962, 1964; Wickelgren, 1965). (Digits presented through noise are acoustically less discriminable from each other than digits presented in clear—and may be less well-recalled for this reason alone.) The alternative hypothesis is the one which we wish to test. The process of recognizing digits through noise may pre-empt channel capacity necessary for their efficient retention in immediate memory storage.

The second hypothesis would be confirmed if the probability of recall of digits presented in clear were reduced when other digits, in the same list, were presented through noise. Such an effect could be either proactive or retroactive. Experiment 2 was therefore designed to test for both these possibilities.

EXPERIMENT 2

Method

Fifty-six lists, each containing eight different digits, were made up by selection from random number tables. Each list was recorded by a male speaker as two groups of four digits read at 1 digit/sec. with a 2 sec. pause between groups and a 10 sec. pause between lists. Subjects were instructed to listen and to remember both groups of digits, but were only required to repeat one group. The group to be recalled was post-cued by the instruction "Group 1" or "Group 2" recorded 1·5 sec. after the end of each list.

The set of 56 lists was then re-recorded through the simulated G.P.O. telephone line and M.N.R.U. described for Experiment 1. The M.N.R.U. was either switched in (Noise) or out of link (Clear). Half-lists were recorded either through noise or in clear to give equal numbers (i.e. 14) with each of the four possible compositions Noise/Noise; Noise/Clear; Clear/Noise and Clear/Clear. Post-cuing instructions were so allocated that the subject was required to recall Group 1 and Group 2 seven times under each of these four conditions of list-composition. Eight conditions of presentation and recall could thus be compared. The order in which subjects experienced list-composition and cueing instructions was subjectively random. All subjects experienced the sequence of lists in the same order.

The set of 56 lists was presented to (89) subjects tested in groups of 11 to 21 people. Conditions were as described for Experiment 1.

Subjects

Subjects were 89 members of the A.P.R.U. civilian panel aged from 25 to 69 years (Mean = 45·3 years).

Results

Half-lists were scored correct if the four digits requested were reported in their correct order. The mean number of half-lists correctly reported in each of the eight presentation and recall conditions are set out in Table II.

TABLE II

MEAN NUMBER OF HALF-LISTS OF DIGITS CORRECTLY REPRODUCED UNDER EIGHT CONDITIONS OF PRESENTATION AND RECALL

EXPERIMENT 2

	Recall of first group				*Recall of second group*			
	C/C	*C/N*	*N/N*	*N/C*	*C/C*	*C/N*	*N/N*	*N/C*
Mean No. of groups correctly recalled	4·53	4·14	3·54	4·44	6·10	5·67	5·91	6·31
σ	3·46	3·8	3·01	2·97	1·66	2·76	2·18	0·13

Analysis of variance gave significant terms for comparisons between presentations in Noise and in Clear ($p<0{\cdot}001$) and for recall of second half-lists as against first half-lists ($p<0{\cdot}001$). The term for differences between subjects was also significant ($p<0{\cdot}01$).

The interesting question was whether the recall of digits in one part of the list was affected if digits elsewhere in the list were presented through noise. This was examined by calculating S^2 from the residual term for the An.o.Va. in order to obtain a rank-order of significant differences for half-lists in each condition of presentation and recall. It appeared that first half-lists (whether presented in noise or in clear) were better recalled when the subsequent half-list was presented in clear than when it was presented in noise ($p<0.01$). The opposite effect was not obtained, that is, recall of a second half-list was no better when the first half-list was presented in clear than when it was presented through noise.

Discussion

The first group of digits was better recalled if the second group was presented in Clear than if it was presented through Noise. There was no evidence that presentation of the first group through Noise affected recognition or retention of the second group. These results cannot be explained on the hypothesis that noise makes digits harder to remember only because it renders them acoustically less discriminable from each other. Consistent with these results is the hypothesis that rehearsal of the first group of digits is inhibited if the channel is pre-empted by the task of recognizing a second group through noise. There are theoretical (Crossman, 1955; Laming, 1968) and empirical (Rabbitt, 1963) reasons for supposing that human channel capacity can be defined (interchangeably) in terms of the number of items between which subjects may have to discriminate (information load) and the mutual discriminability of the items with which subjects are concerned. This result is therefore consistent with Posner and Rossman's (1965) evidence that the maintenance of items in immediate memory requires channel-capacity which may be pre-empted by secondary tasks in proportion to their information load. It further appears that the recognition of digits through noise constitutes such a "secondary" task, interfering with the retention of digits previously heard. Items presented through noise may be poorly remembered (Exp. 1) not only because they are incorrectly recognized, but because the difficulty of recognizing some items interferes with the retention of others (Exp. 2).

It is rare that the act of "communication" for which a speech-link is used simply involves an exchange of lists of words (Rabbitt, 1966) or of groups of digits (Exp. 1 and 2). The slow tempo of conversation and the considerable redundancy of language make it possible that under everyday circumstances the burden on a listener's information processing system would be low enough to allow him sufficient time (and so, implicitly, sufficient channel-capacity) to both understand and remember the content of continuous speech transmitted to him through low levels of noise. A more naturalistic experiment, therefore, would test whether recall of the content of connected discourse presented in clear is affected when, before recalling it, a listener is subsequently required to try to understand and remember speech presented to him through noise. An experiment was made to test this point.

EXPERIMENT 3

Apparatus and procedure

Two prose passages, 682 and 712 words long, were made up from the text of articles in back numbers of "Scientific American." These were then recorded on magnetic tape by a male speaker at approximately 120 words/min. The durations of the passages, as read, were 5 min. 8 sec. (Passage A) 5 min. 23 sec. (Passage B).

The master tape on which these passages were recorded was then re-recorded through a simulated G.P.O. telephone link and G.P.O. Modulated Noise Reference Unit (M.N.R.U.).

Each of the two passages was re-recorded once with the M.N.R.U. switched out of the circuit (No Noise Condition) and once with the M.N.R.U. switched out for the first half of the passage, and switched in for the second half at a setting which maintained the noise intensity 5 dB. below the intensity of the speech signal (Clear/Noise Condition). The halves of the passage were defined with respect to the sentence-ending nearest to the mid-point located by a word-count.

For each passage a list of 10 questions was made up to test the listener's recall of the material read to him. The first five questions could be answered only by reference to information presented in the first half of the passage. No information relevant to these questions was repeated in the second half of the passage. The second five questions similarly required information presented only in the second half of each passage.

Separate groups of 11 to 22 subjects each heard one of these passages recorded under one of these conditions (i.e. No Noise, or Clear/Noise). Passages were played to subjects on a "Ferrograph" 5A tape-recorder through a Phillips ZA 29431 external speaker system.

The Clear/Noise condition, when used, was described to the subjects in advance. The passage was then played. After the passage had ended subjects removed appropriate question sheets from envelopes and wrote down their answers to the questions. In this way 36 subjects heard passage A in the No Noise Condition and 36 heard passage A in the Clear/Noise Condition. Similarly two groups of 26 subjects each heard passage B under one of these conditions.

Subjects

The subjects were 124 members of the A.P.R.U. civilian subject panel. Thirty-eight of these people were men and 86 were women. Their ages ranged from 22 to 68 years (Mean = 42·6 years). Fifty-two of these subjects had previously participated in experiments investigating long- and short-term memory.

Results

Questions were framed so that answers would be unambiguously correct or wrong; i.e. a single name, figure or binary (Yes/No) choice was required by each. The scorer awarded 1 point for each correct answer, and 0 points for any other answer. The mean scores made by subjects on the first and second halves of each passage in each condition are given in Table III below.

TABLE III

MEAN NUMBERS OF QUESTIONS CORRECTLY ANSWERED

EXPERIMENT 3

	N	*First half of passage*	*Second half of passage*
Passage A			
No Noise	36	2·1 ($\sigma = 1·7$)	3·2 ($\sigma = 1·9$)
Clear/Noise	36	1·7 ($\sigma = 1·4$)	2·5 ($\sigma = 1·6$)
Passage B			
No Noise..	26	1·8 ($\sigma = 1·6$)	2·6 ($\sigma = 1·5$)
Clear/Noise	26	1·2 ($\sigma = 1·1$)	2·4 ($\sigma = 1·8$)

Separate groups of subjects heard each of two passages under two conditions. Scores represent the mean number of questions relating to first and second halves of passages correctly answered (max. possible 5 in each case).

For each passage scores on the first five questions obtained by the No Noise group were compared by a Mann-Whitney U-test against equivalent scores obtained by the Clear/Noise group. For both passages, scores obtained by the No Noise group were significantly higher (Passage A, $p<0·05$ two-tailed and Passage B, $p<0·025$ two-tailed). Recall scores for the second halves of the passages were also significantly higher for the No Noise groups than for the Clear/Noise groups ($p<0·01$. two-tailed, in each case). These latter differences do not concern the present argument, since we cannot tell whether they mean that subjects did not correctly hear, or whether they could not remember the material presented to them. There was no interaction between the success with which individual questions were answered, and the effects of noise.

Discussion

This experiment suggests a complex picture of the activity of a subject who attempts to remember a prose passage which is read aloud to him. Attention to a continuous flow of new data reduces the opportunity for rehearsal of material already in store and renders its recall more uncertain. A skilled listener must therefore, presumably, learn to divide his attention between the analysis of new information and the rehearsal of information in store. In doing this he will be considerably aided by redundancy (either in stored material or in current input). It is well known that the recall of prose passages has the character of a précis, in which certain salient details are related to each other in terms of the system of causal relationships which are most probable (and so redundant) in the context of the subject's life-experience (Bartlett, 1932). It is tempting, therefore, to say that the subject recognizes redundancy in prose passages, and does not tax his limited capacity by attempting to store redundant material verbatim, since he can reconstruct redundant connecting links from data in long-term memory storage ("schemata").

This usage conceals an assumption about the mechanism necessary to recognize and use redundancy. It may not be evident to a listener which details are "salient" and which "redundant" until much of a passage has been heard. Evidently a listener would operate most efficiently if he did not rehearse all the material presented to him, as it were by rote, but rather attempted selective rehearsal, editing material in storage both in terms of the new information which he continuously receives and in terms of other information which he may hold in long-term memory. Thus the "salient" and "redundant" features of a passage may be progressively distinguished and selectively rehearsed or forgotten as the episode progresses.

Such a process would evidently require an highly complex and carefully-timed selective shifting of attention between material currently being presented and material held in storage. It would be consistent with this model that any increase in the complexity of one of these interlocking and competing processes would reduce the efficiency with which the others can be carried on. In this context it is again evident that under conditions approximating to those encountered in everyday life levels of noise on speech-communications links which do not prevent listeners from hearing what is said to them may nevertheless prevent them from efficiently using the information which they receive.

Apart from these theoretical considerations the present results raise a clear practical issue. Intelligibility tests are again shown to provide very crude indices of the extent to which noise on speech communications links may limit the efficiency of the people who use them (Broadbent, 1958; Rabbitt, 1966). Communications links, such as telephones, are not simply used to transmit lists of words or phonemes but often to conduct complex transactions. Because such transactions make demands on memory storage and retrieval, any attenuation of these functions may represent a serious liability. In assessing the efficiency and comfort with which systems may be used it is evidently necessary to bear in mind that the effects of noise cannot be disregarded until a point is reached at which recognition errors occur. The present results suggest that even levels of noise which do not have measurable effects on intelligibility may nevertheless cause measurable decrements in the ability of listeners to carry out the transactions for which communications systems may be required.

REFERENCES

BARTLETT, F. C. (1932). *Remembering.* London: Cambridge University Press.

BROADBENT, D. E. (1958). *Perception and Communication.* London: Pergamon. Pp. 73–80.

BROWN, I. D. (1964). The measurement of perceptual load and reserve capacity. *Trans. Ass. industr, med. Offrs.* **14**, 44–9.

CHISTOVITCH, L. A. (1965). In CHISTOVITCH, L. A. and KOZHEVNIKOV, V. A. (Eds.). *Rech 'Artikulatsia i Vospriatie.* Moscow.

CONRAD, R. (1962). An association between memory errors and errors due to acoustic masking of speech. *Nature,* **193**, 1314–5.

CONRAD, R. (1964). Acoustic confusions in immediate memory. *Brit. J. Psychol.*, **55**, 75–84.

CROSSMAN, E. R. F. W. (1955). The measurement of discriminability. *Quart. J. exp. Psychol.*, **7**, 176–95.

KRYTER, K. D. (1960). Human engineering principles for the design of speech communications systems. *Operations Applications Office, Air Force Command and Control Development Centre, A.F.C.C.D.D. Tech. Report* 60. 27.

LAMING, D. R. J. (1968). *Information Theory of Choice-Reaction Times.* London: Academic Press.

POSNER, M. I., and ROSSMAN, E. (1965). Effect of size and location of informational transforms upon short-term retention. *J. exp. Psychol.*, **70**, 496–505.

POULTON, E. C. (1958). Measuring the order of difficulty of visual-motor tasks. *Ergonomics*, **1**, 234–9.

RABBITT, P. M. A. (1963). Information load and discriminability. *Nature.* **197**, 726.

RABBITT, P. M. A. (1966). Recognition: memory for words correctly heard in noise. *Psychon. Sci.*, **6**, 383–4.

WICKELGREN, W. A. (1965). Acoustic similarity and intrusion errors in short-term memory. *J. exp. Psychol.*, **70**, 102–8.

WEBSTER, J. C., and KLUMPP, R. C. (1965). Speech interference aspects of Navy noises. *Res. Develop. Rep. NEL/REP.* 1314. *U.S. Navy Electronics Lab., San Diego, California.*

Manuscript received 12th February, 1968.

HOW DOES ACOUSTIC SIMILARITY INFLUENCE SHORT-TERM MEMORY?

BY

A. D. BADDELEY*

From the Medical Research Council, Applied Psychology Research Unit, Cambridge

This study attempts to discover why items which are similar in sound are hard to recall in a short-term memory situation. The input, storage, and retrieval stages of the memory system are examined separately. Experiments I, II and III use a modification of the Peterson and Peterson technique to plot short-term forgetting curves for sequences of acoustically similar and control words. If acoustically similar sequences are *stored* less efficiently, they should be forgotten more rapidly. All three experiments show a parallel rate of forgetting for acoustically similar and control sequences, suggesting that the acoustic similarity effect does not occur during storage. Two input hypotheses are then examined, one involving a simple sensory trace, the other an overloading of a system which must both discriminate and memorize at the same time. Both predict that short-term memory for *spoken* word sequences should deteriorate when the level of background noise is increased. Subjects performed both a listening test and a memory test in which they attempted to recall sequences of five words. Noise impaired performance on the listening test but had no significant effect on retention, thus supporting neither of the input hypotheses. The final experiments studied two retrieval hypotheses. The first of these, Wickelgren's phonemic-associative hypotheses attributes the acoustic similarity effect to inter-item associations. It predicts that, when sequences comprising a mixture of similar and dissimilar items are recalled, errors should *follow* acoustically similar items. The second hypothesis attributes the effect to the overloading of retrieval cues which consequently do not discriminate adequately among available responses. It predicts maximum error rate *on*, not following, similar items. Two experiments were performed, one involving recall of visually presented letter sequences, the other of auditorily presented word sequences. Both showed a marked tendency for errors to coincide with acoustically similar items, as the second hypothesis would predict. It is suggested that the acoustic similarity effect occurs at retrieval and is due to the overloading of retrieval cues.

INTRODUCTION

One of the most striking aspects of short-term memory for sequences is its extreme vulnerability to the effects of acoustic similarity among the items concerned. This was first demonstrated by Conrad and Hull (1964) who showed that sequences of letters that are hard to discriminate when heard in noise (e.g. CBGPTV) are retained much less accurately than easily discriminable sequences (e.g. RXQHKW), even though they are presented visually. This effect has also been shown to apply to the recall of consonant-vowel digrams (Wickelgren, 1965*a*) and words (Conrad, 1963), for which Baddeley (1966*a*) has shown it to be considerably more important than either semantic or formal similarity. In this it differs from long-term memory which appears to rely more on semantic than acoustic coding (Baddeley, 1966*b*; Baddeley and Dale, 1966).

Since the acoustic similarity effect is such a marked and characteristic phenomenon of short-term memory, it seems reasonable to suppose that an adequate explanation would have considerable implications for a general model of short-term memory. Unfortunately, however, the facts as they stand can be explained equally well by a decay theory (Conrad, 1964), an interference theory (Wickelgren, 1965*b*) or, with a little ingenuity, by any of several models based on an informational approach to short-term memory of the type described by Broadbent (1958), Posner and Rossman (1965) and Murdock (1965). The present study therefore does not attempt to test existing theories directly; its aim is rather to narrow down the range of possible theories by first asking more general questions.

---

* Now at The Laboratory of Experimental Psychology, University of Sussex.

Any memory system must have at least three components; it must have a means of storing information, a system for putting information into the store, and a mechanism for retrieving it from the store when recall is required. The present study is concerned with two questions; which of these stages is responsible for the acoustic similarity effect, and what sort of mechanism is involved?

I. Storage

Both the decay and interference theory explanations of the acoustic similarity effect have implied a model in which sequences are encoded acoustically and are forgotten because the cues required to discriminate between items are lost during storage, either because of spontaneous decay (Conrad, 1964) or interference among the items stored (Wickelgren, 1965*b*). Since similar items have few discriminating cues they should be more drastically affected by any loss that occurs during storage.

The experiments which follow are therefore concerned with the question of whether the decrement due to acoustic similarity does in fact occur during the storage phase. The technique used is based on that used by Brown (1958) and Peterson and Peterson (1959) in which the subject is presented with one or more items for retention. He is then given a second task such as counting backwards, which prevents his rehearsing between the presentation of the item and the subsequent recall. Using this technique Peterson and Peterson showed a forgetting curve which reached an asymptote after about 20 sec. Since both input and retrieval are held constant, and only storage time is varied, it seems reasonable to assume that the forgetting shown represents loss during storage. If the acoustic similarity effect occurs during storage, then sequences of acoustically similar items should show a steeper forgetting curve than control sequences.

EXPERIMENT I

Material. Fifteen sequences of four words were drawn at random from each of two sets, one comprising 10 acoustically similar words (man, cab, can, cad, cap, mad, max, cat, map), and the other comprising 10 control words of equivalent frequency of occurrence in the language (few, pit, cow, pen, sup, bar, day, hot, rig, bun).

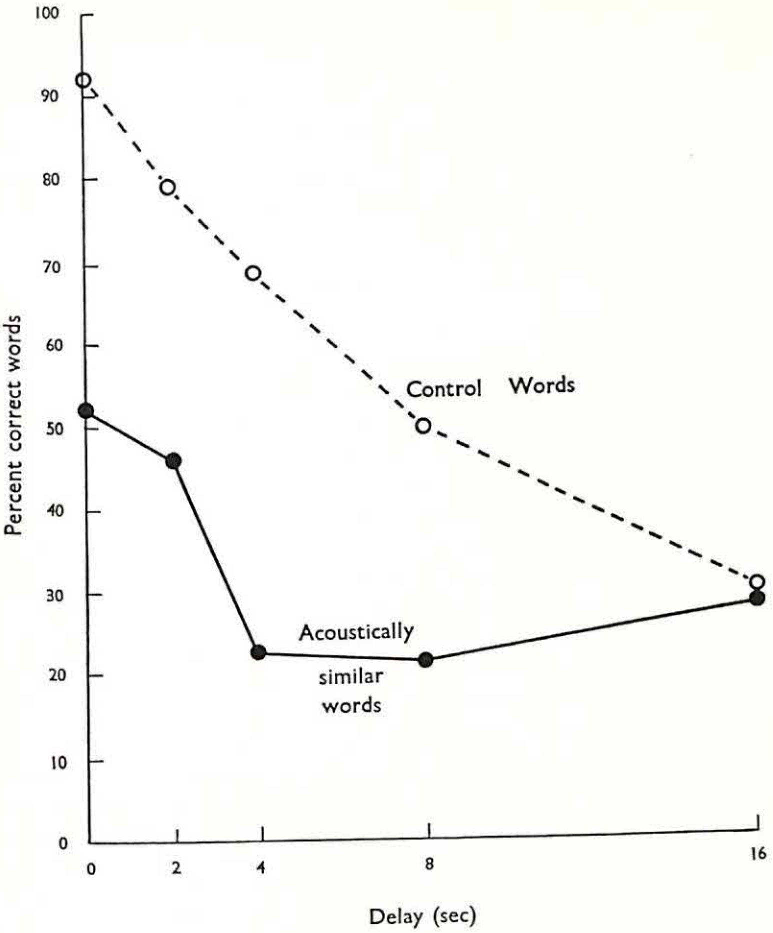
Procedure. Word sequences were spoken at a rate of one word per sec. and were tested after a retention interval of 0, 2, 4, 8 and 16 sec. during which subjects copied down random digits also spoken at a one sec. rate. After the requisite number of digits a verbal recall signal was given, whereupon subjects were allowed 12 sec. to write down the word sequence in the order presented. Both digits and words were written on the same line of the response sheet. Thus the location of the words was determined by the number of digits to be copied and, since subjects did not know this in advance, they could not "jump the gun" by interleaving the recall and digit copying tasks. In order to maximize response availability, the two sets of words were written on cards which were permanently visible. The main experiment was preceded by six practice items, one from each word set with successive retention intervals of 0, 4 and 8 sec. Subjects were encouraged but not compelled to guess when in doubt.

Subjects were 15 housewives from the A.P.R.U. subject panel, they were tested in a single group. Order of occurrence of word sequences and retention intervals was randomized.

Results

Performance was scored in terms of mean per cent. words correctly reproduced in the appropriate serial position and is shown in Figure 1. It is clear from this that

FIGURE 1



Rate of forgetting sequences of four acoustically similar or control words.

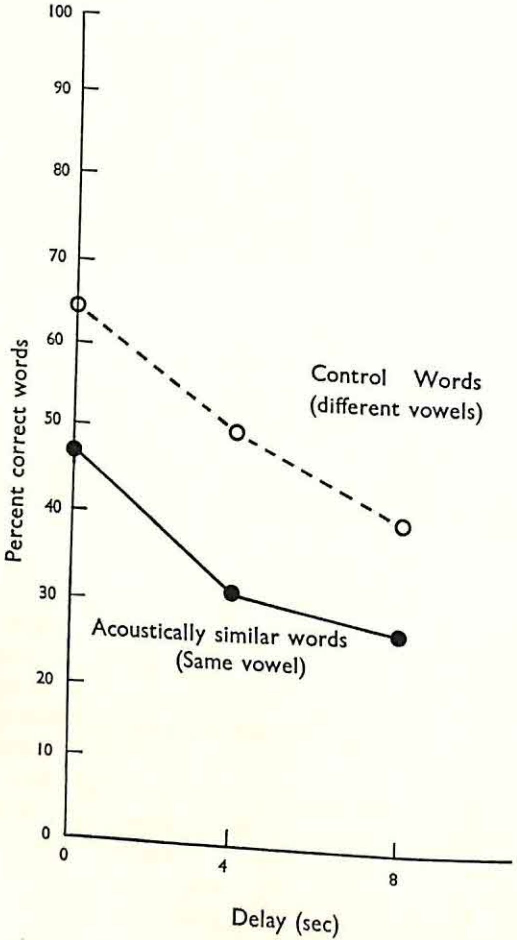
sequences of both types show the rapid forgetting demonstrated by Peterson and Peterson (1959). It is also clear that, although the acoustically similar sequences were considerably harder than the controls for all subjects, the differential rate of forgetting predicted by the storage hypothesis does not occur. Indeed, considered as a whole, the interaction between type of material and rate of forgetting is in the opposite direction to that predicted, with a drop of about 30 per cent. for acoustically similar sequences and 60 per cent. for controls. However, such a comparison would clearly be invalid as an indicator of rate of short-term forgetting since the acoustically similar sequences start at a much lower level of performance and are clearly approaching an asymptote by 4 sec. Control sequences start at a higher level and forgetting continues up to about 16 sec. Even comparison over the first 4 sec. however, is not likely to be valid since the initial point for control sequences is almost certainly somewhat depressed due to a "ceiling" effect. In short, Experiment I does not show a difference in rate of forgetting of any magnitude between the two types of material, but we can not discount the possibility that an effect exists but is masked by the difference in initial level of performance. Experiment II attempts to avoid this problem.

EXPERIMENT II

Material. This was selected with two aims in view, first to reduce the difference in difficulty between similar and control sequences so as to avoid the artefactual "ceiling" and "floor" effects outlined above, and secondly to minimize the possibility of using different coding systems for the two conditions. It was, for example, possible with the previous materials to remember only the initial consonants of control sequences and subsequently reconstitute the sequence from the permanently visible prompt card, a strategy which is clearly not practicable with the acoustically similar sequences. The two sets of words in Experiment II were therefore matched in terms of initial and terminal consonants and differed only in the range of vowels used, with the acoustically similar set comprising 10 words with the same vowel (had, tag, fad, cap, pad, pan, tap, lap, fan and can) and the control list containing all five vowels (tin, den, cod, lid, cab, tug, fog, dig, fun, led).

Procedure. This was as in Experiment I except that (1) Presentation was visual, the words being written on cards which were exposed at a rate of one per sec. (2) Sixteen sec. was allowed for recall. (3) Four sequences from each set were tested with each of three retention intervals, 0, 4 and 8 sec., making a total of 24 sequences. (4) The order of presentation of the 24 sequences was randomized. A separate order was used for each of two groups of 30 young servicemen.

FIGURE 2



Rate of forgetting visually presented four-word sequences containing the same or different vowels.

Results

Performance on the two types of material is shown in Figure 2. As was intended, the difference in initial difficulty between the two sets was smaller than the 40 per cent. shown in Experiment I, mainly due to the greater difficulty of the control sequences and presumably caused by the greater acoustic similarity resulting from duplication among both initial and terminal consonants. Nevertheless, there was a highly significant difference in difficulty between the two sets of words, an effect shown by all but five of the 60 subjects ($p < 0{\cdot}001$, Sign Test). Rate of forgetting appears to be remarkably similar for the two sets of words, both overall, and over the first 4 sec. Comparison of rate of forgetting in the two conditions using the Wilcoxon Test confirmed this conclusion ($p > 0{\cdot}1$ in each case).

This result then agrees with Experiment I in suggesting that the acoustic similarity effect is not due to the storage phase of short-term memory. However, in both experiments, control sequences start at a higher level of performance, and it might be argued that if subsequent retention were measured as a percentage of initial performance, then similar sequences do show more forgetting than control sequences. While such a procedure might seem somewhat dubious, it would clearly be desirable to repeat these experiments with level of intial performance equated for acoustically similar and control sequences. One method of achieving this is to have acoustically similar and control sequences of different length, the relative lengths chosen so as to give the same per cent. correct score when tested immediately. This technique was used in Experiment III.

EXPERIMENT III

Method. The design employed was identical with Experiment I except that acoustically similar sequences comprised three words and control sequences five words, lengths which previous experimentation had suggested would give a score of approximately 80 per cent. correct on immediate test. The subjects, 17 housewives from the A.P.R.U. panel were tested in a single group.

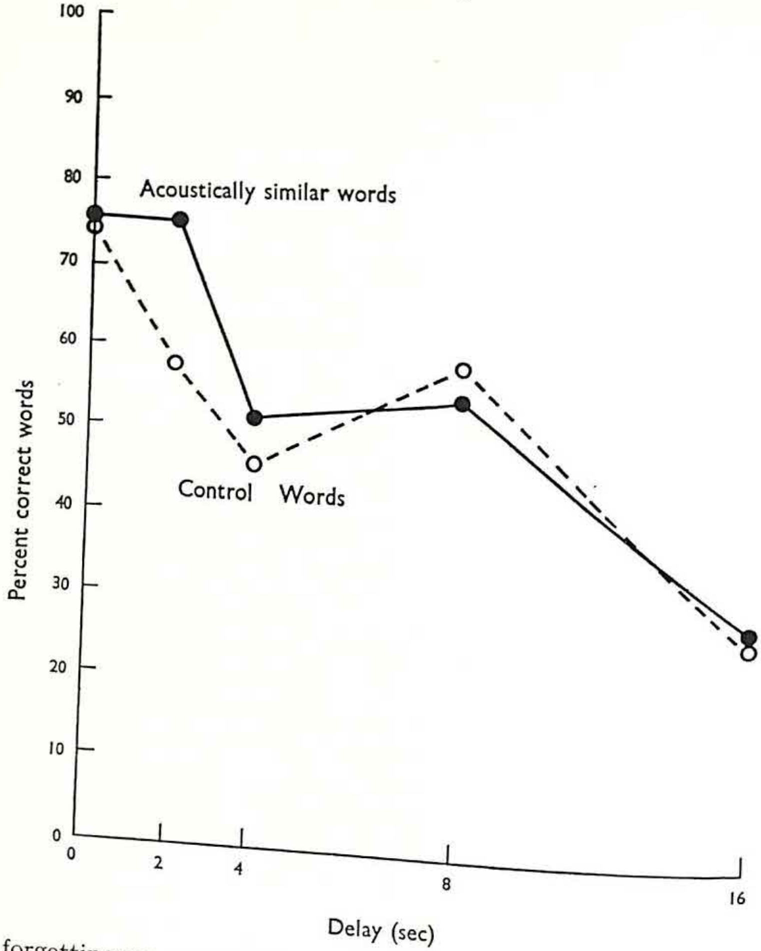
Results

The percentage of words reproduced correctly in the appropriate serial position is shown in Figure 3. Level of difficulty of acoustically similar and control sequences on immediate test appears to have been successfully equated. As in Experiments I and II there is no indication whatsoever that the acoustically similar sequences are forgotten more rapidly than control sequences, as the storage hypothesis would predict.

DISCUSSION

All three experiments suggest that rate of short-term forgetting is not affected by acoustic similarity. They also suggest that rate of forgetting is independent of both initial level of performance, and of number of items to be retained. Experiments I and II showed no difference in rate of forgetting between difficult acoustically similar sequences and relatively easy control sequences. This contrasts with the evidence from L.T.M. which suggests that the higher the level of initial performance the better the retention (Underwood, 1949). Furthermore the observation in Experiment III that three- and five-word sequences were forgotten at the same rate contradicts the theoretical short-term forgetting function proposed by Melton (1963) which postulates a large interaction between number of items and rate of forgetting. Melton's generalization was based principally on an S.T.M. study in which subjects were required to retain sequences of one, two, three, four and five consonants for

FIGURE 3



Rate of forgetting sequences of three acoustically similar or five control words.

periods of 4, 12 or 32 sec. The interaction observed depended on (1) the fairly plausible assumption that if tested immediately subjects could have reproduced any of the sequences perfectly, i.e. all the curves originated at the 100 per cent. correct point, and (2) on subsequent tests per cent. correct varied systematically with number of consonants to be retained, so that after 32 sec. over 90 per cent. of the single consonant items were correct, compared with less than 10 per cent. of the five consonant sequences. It is, however, only reasonable to accept this interaction as a valid indication of real differences in rate of forgetting if one assumes that the 100 per cent. correct point represents the same amount of learning for the one and five-consonant sequences, a very implausible assumption which could no doubt be tested by using some more sensitive test of degree of learning such as response latency. In other words, at points near 100 per cent. correct, performance is unlikely to give an accurate indication of the level of underlying learning or retention. If points which are likely to be influenced by a "ceiling" effect of this sort are discounted, the striking feature of Melton's results is how *parallel* is the rate of forgetting for sequences of different length. As Melton suggests, the relationship between length of sequence and rate of short-term forgetting has important implications for any theory of S.T.M. It would therefore seem desirable to repeat Melton's study taking possible scalar artefacts into account.

To return to the acoustic similarity effect, however, the three experiments described all indicate a parallel rate of forgetting for acoustically similar and control sequences. It is difficult to see how these results can be explained by a theory attributing the acoustic similarity effect to the storage phase of short-term memory (Conrad, 1964; Wickelgren, 1966).

II. The Input Stage

The next experiment is concerned with two classes of hypothesis which would attribute the acoustic similarity effect to the input phase. The first of these could be termed a simple acoustic trace hypothesis. It has been shown that subjects tend to remember sequences of unrelated words in terms of their sounds unless they are presented several times (Baddeley, 1966*a*, 1966*b*). It therefore seems plausible to assume that in these circumstances the subject simply records the raw sound as a primary acoustic memory trace (cf. Mackworth, 1964). Items which are acoustically similar will thus lay down similar traces which will be difficult to discriminate and hence to recall. A hypothesis of this type has been described by Murdock (1966) who suggests that S.T.M. may draw on both a memory store and a pre-perceptual sensory store of this sort.

A second, more complex hypothesis however, might explain the acoustic similarity effect as follows. The role of the input stage of S.T.M. is to discriminate the message to be remembered, encode it as efficiently as possible and then to feed it into the memory store. Each of these processes will require a certain amount of the limited channel capacity available (Posner and Rossman, 1965; Murdock, 1965). If the informational demands of these processes is increased, then the channel will become overloaded and retention will suffer. Since material is encoded acoustically, acoustically similar items will overload the system by presenting the subject with a series of difficult auditory discriminations (whether of external or, when presentation is visual, of internal self-generated stimuli). Since more channel capacity is being used for discrimination, less will be available for encoding, rehearsal, etc., and the level of performance will consequently decline.

On either of the above hypotheses, S.T.M. should suffer when the words to be retained are presented auditorily in noise. On the simple acoustic trace hypothesis, a word spoken in noise should leave a trace in which the word is embedded in noise, and as a result retrieval should be impaired. On the channel capacity hypothesis, the need to make a series of relatively difficult discriminations in order to hear the message should reduce the capacity available for encoding and rehearsal, with a consequent loss in amount recalled.

Evidence already exists that S.T.M. for auditory material may be impaired by noise, thus Rabbitt (1966) has shown that noise may impair recognition memory for auditorily presented words, though the effect shows up as higher false alarm rate rather than a lower detection score. Similarly Murdock (1967) has shown that minimal paired-associate learning is poorer for pairs of words presented in noise. The question at issue, however, is not whether noise has an effect on S.T.M., but whether its effect is large enough to support an explanation of the massive acoustic similarity effect in terms of stimulus discriminability.

EXPERIMENT IV

Material. Twenty-four sequences of five words were drawn from each of two sets, an acoustically similar set (man, mad, cad, map, can, cap, cat, mat) and a control set of equivalent frequency of occurrence in the language (sup, day, pit, hot, pen, bar, cow, few). Sequences were drawn at random with the constraint that words never occurred more than once within the same sequence.

Procedure. The sequences were presented by tape recorder at a rate of one word per sec. after which subjects were allowed 20 sec. to write down as much of the sequence as they could recall in the appropriate order. The experiment was divided into two halves, each involving the presentation of twelve sequences from each set in random order. In one half all sequences were presented against a background of 70 dB. white noise, while in the other half the background was relatively quiet (45 dB.). At each noise level the remembering task was preceded and followed by a listening test in which all 16 words were read out in random order. Subjects were allowed 5 sec. per word to write their responses. Combined speech and background noise level was 70 dB. in the quiet condition and 74 dB. in the noisy condition. Subjects were housewives from the A.P.R.U. volunteer subject panel; they were tested in two groups, one ($N = 20$) was given the noisy condition first and the other ($N = 18$) did it second. To maximize response availability, the two sets of words were displayed on cards throughout the experiment, though subjects did not know in advance from which set the next sequence would be drawn.

Results

Listening test. Table I shows the per cent. words correctly reported for the two sets at the two levels of background noise. The 70 dB. background noise clearly impared discrimination of similar words ($p < 0{\cdot}001$, Sign Test), and although its effect on control words was small, it was consistent and statistically reliable ($p < 0{\cdot}02$, Sign Test). Acoustically similar words were harder to discriminate than control words both in quiet ($p < 0{\cdot}01$ Sign Test) and noise ($p < 0{\cdot}001$, Sign Test). Control words in noise had the same level of discriminability as acoustically similar words in quiet, seven subjects found them easier, seven harder and 24 equally discriminable.

TABLE I

Percent Words Correct on Listening Test

Acoustically similar words				*Control words*			
	Order of test				*Order of test*		
Condition	*1st*	*2nd*	*Mean*	*Condition*	*1st*	*2nd*	*Mean*
Noise	78·4	92·0	85·2	Noise	97·8	99·7	98·8
Quiet	97·6	99·9	98·8	Quiet	99·7	99·9	99·8
Mean	88·0	96·0		Mean	98·8	99·8	

Recall. Table II shows per cent. words recalled in the correct position for the various conditions. As expected, acoustically similar sequences are harder than controls, an effect that is shown by all 38 subjects at both noise levels. The effects of noise on recall are slight, unreliable and do not approach statistical significance. It may seem paradoxical that a difference due to noise of 1 per cent. in the listening test for control words should be significant whereas a difference of 5·1 per cent. in recall score should not ($T = 237{\cdot}5$, $Z = 1{\cdot}03$, $p > 0{\cdot}1$). The reason for this is that whereas only one subject performed better on the control word *listening* task in noise, 15 subjects *recalled* control words better in noise. Indeed, with acoustically similar words the effect was even more marked, with 21 of the 38 subjects having a higher recall score in noise, probably because of practice effects. Poulton and Freeman (1966) have shown that a counterbalanced design of this sort is particularly vulnerable

TABLE II

Per Cent. Words Correctly Recalled

Acoustically similar words				*Control words*			
	Order of test				*Order of test*		
Condition	*1st*	*2nd*	*Mean*	*Condition*	*1st*	*2nd*	*Mean*
Noise	43·4 (38·2)	52·6 (49·4)	48·0 (43·8)	Noise	84·7 (90·7)	90·3 (92·5)	87·5 (91·6)
Quiet	47·6	53·6	50·6	Quiet	92·5	92·7	92·6
Mean	45·5	53·1		Mean	88·6	91·5	

Figures in parentheses represent scores expected in noise on the basis of the listening test (see text).

to asymmetrical transfer effects which may either produce a spurious effect or mask a real one. This was checked by making a comparison across groups using only the first condition tested in each case. This supported the previous analysis; the Mann Whitney Test showed that the difference between recall of words presented in quiet and in noise did not approach statistical significance for either acoustically similar or control words ($p > 0{\cdot}1$ in both cases). Finally, there was a very reliable effect of practice, with 28 of the 38 subjects improving their score on similar sequences and 26 improving on control words ($p < 0{\cdot}01$, Sign Test in each case). This small but reliable practice effect appears to have been sufficient to overshadow any effect that noise might have had on recall.

The results of the listening test show that 70 dB. of white noise impaired discrimination for both acoustically similar and control words. The retention test, however, indicates that noise did not reliably impair recall. This is particularly striking since comparison between quiet and noise was made using raw scores which should represent not only recall errors, but also errors due to mishearing. The listening tests scores may be used to correct these raw scores as follows. If, for example, noise reduces the listening scores from 100 to 90 per cent., it seems reasonable to suppose that 10 per cent. of the words to be recalled will be misheard, so that they will be wrong even though no forgetting occurs. On this basis the expected raw recall score would be 90 per cent. of that in quiet. Recall scores expected in noise on the basis of recall in quiet and listening in noise are shown in parentheses in Table II. They reinforce the conclusion based on the analysis of raw scores in suggesting that presenting word sequences in noise had little effect on their retention. This does not mean that S.T.M. is insensitive to noise; the results of Rabbitt (1966) and Murdock (1967) show that impairment due to noise may occur under appropriate conditions. It does however suggest that any effect under these conditions is small enough to be masked by other factors.

Discussion

Since these results show no impairment of S.T.M. in noise, they support neither of the two input-stage hypotheses of the acoustic similarity effect. Control sequences in noise and acoustically similar sequences in quiet are equally difficult to hear, but differ enormously in retention score, suggesting that subjects are able to filter out the noise before storing the stimulus. Indeed, it seems possible (Hintzman, 1965) that the stimulus is coded as a kinaesthetic vocal image, rather than acoustically, in which case the noise would almost inevitably be filtered out and discarded. It seems, moreover, that this noise filtering process requires relatively little channel capacity, for there is abundant evidence that short-term memory functions as a system of limited informational capacity (Broadbent, 1958), so that any concurrent information processing required will adversely affect recall performance (Posner and Rossman, 1965; Murdock, 1965).

This experiment does not of course completely eliminate the input phase as a possible source of the acoustic similarity effect. There may for example be a further stage in which the filtered stimulus is registered in the memory store by an encoding process of limited capacity, at which point the adequate encoding of similar material may overload the system. If this were the case, however the acoustic similarity effect should become more acute as rate of presentation is increased. Experiments by Conrad, Baddeley and Hull (1966) and Laughery and Pinkus (1966) have shown no such interaction between the acoustic similarity effect and rate of presentation for speeds ranging between 20 and 180 items/min. As an explanation of the massive acoustic similarity effect, therefore, the hypothesis of limited input capacity does not seem very probable.

The results of this study have implications at two levels. At the practical level, the absence of a reliable effect of listening conditions on S.T.M. suggests that experiments on acoustic similarity may use auditory presentation with little danger of introducing major artefacts. Provided subjects are able to hear the stimuli correctly, auditory and visual presentation of acoustically similar material should and do give substantially the same results (Baddeley, 1966*a*, *b*). At the theoretical level these results indicate that neither of the input hypotheses described can account for the acoustic similarity effect in S.T.M. This, together with the failure of Conrad, Baddeley and Hull (1966) to find an interaction between the acoustic similarity effect and rate of presentation, makes an input phase explanation seem unlikely.

III. *Retrieval*

It is, in principle, possible to generate a wide range of retrieval hypotheses. The following experiments are concerned with two of the possible hypotheses which appear to be both plausible and to produce different predictions.

The first hypothesis was proposed by Wickelgren (1965*b*) who points out that acoustic similarity leads to poorer performance because of failure to reproduce items in the correct *order*, not because the items themselves are less available, indeed recall of the items irrespective of order may be *better* for acoustically similar material. He suggests that subjects remember lists of consonants as sequences of phonemes. Lists containing consonants with a common phoneme, e.g. P and T in X*P*WR*T*J become formally equivalent to lists with repeated items, e.g. 173974, which have been shown to be harder to recall than lists without repeats (Wickelgren, 1965*c*). Wickelgren suggests an associative hypothesis in which each item serves as a stimulus which evokes the next item in the sequence. In the sequence above, the digit 7 serves as a stimulus for both 3 and 4 and thus provides a more ambiguous recall cue than the equivalent digit in a non-repeating sequence. The problem of retrieving a remembered sequence is thus functionally equivalent to that of running through a maze with a choice-point for each unit to be retrieved. A repeated item places the subject as the same choice point on two occasions and requires him to make a different decision on each. The fact that subjects *can* recall repeating sequences successfully clearly implies that association between successive items is not the only cue available; the fact that such sequences are more difficult suggests that such associations are one source of retrieval cues. Acoustically similar letters tend to have phonemes in common and a list of such letters can thus be considered as a sequence containing many repetitions. While this hypothesis would predict that acoustically similar sequences would be more difficult than control sequences, it would not predict faster forgetting, as would a storage hypothesis.

The essence of Wickelgren's retrieval hypothesis is that the subject reproduces acoustically similar material in the wrong order because he loses his place in the sequence. An adequate retrieval process, however, must tell the subject not only where he is in the sequence being retrieved, but also what item to retrieve next. One alternative to Wickelgren's hypothesis might suggest that the subject's difficulty lay in deciding what item to select next rather than where he was in the sequence. On this hypothesis, retrieval is analogous to a treasure hunt in which the subject is required to collect a series of items in a specified order and is given a sequence of clues. The more similar the items, the harder it becomes to decide which clue refers to which item, and the greater the probability of collecting the items in the wrong order. On this hypothesis, retrieval cues serve as an address, specifying for each position what sort of response is appropriate (cf. Crossman, 1959).

There is fortunately one situation in which the Wickelgren and the present address hypotheses make quite different predictions. When a subject is required to retrieve sequences comprising a mixture of similar and dissimilar items, Wickelgren's hypothesis would predict that errors would *follow* similar items as is the case with repeated items (Wickelgren, 1965*c*; 1966). The address hypothesis, however, would predict that the similar items themselves would be wrong, since their addresses are insufficiently discriminable. The following experiments examine this point by requiring subjects to recall mixed lists and examining the error patterns.

EXPERIMENT V

Method. Subjects attempted to recall sequences of six letters drawn from two basic sets, one comprising acoustically similar letters (BCDPTV, probability of acoustic confusion (Conrad, 1964) = 0·699), the other, dissimilar letters (JKLRWY, probability of acoustic confusion = 0·193).

Six basic types of sequence were produced with the constraints that no letter occurred twice in any sequence, and each letter appeared equally often at each serial position. The six arrangements are given below with examples; S represents a letter from the acoustically similar set (BCDPTV) and D represents a letter from the dissimilar set (JKLRWY).

Type	*Arrangement*	*Example*
A	DSDSDS	JCWPLD
B	SDSDSD	CJPWDL
C	SSSDDD	CPDJWL
D	DDDSSS	JWLCPD
E	DDDDDD	JWLYRK
F	SSSSSS	CPDVBT

The test comprised 24 sequences, four of each type. A film was made up comprising three practice sequences followed by the 24 test sequences in random order. A Specto frame-by-frame projector was used to present the letters singly at a rate of 0·75 sec. per letter. Each sequence was preceded by a "ready" dot exposed for 2·25 sec., and recall was unpaced: projection stopped until all subjects had ceased writing. Subjects were instructed to recall in the order of presentation and to guess rather than leave blanks. To maximize response availability, the 12 relevant letters were written at the top of the response sheets.

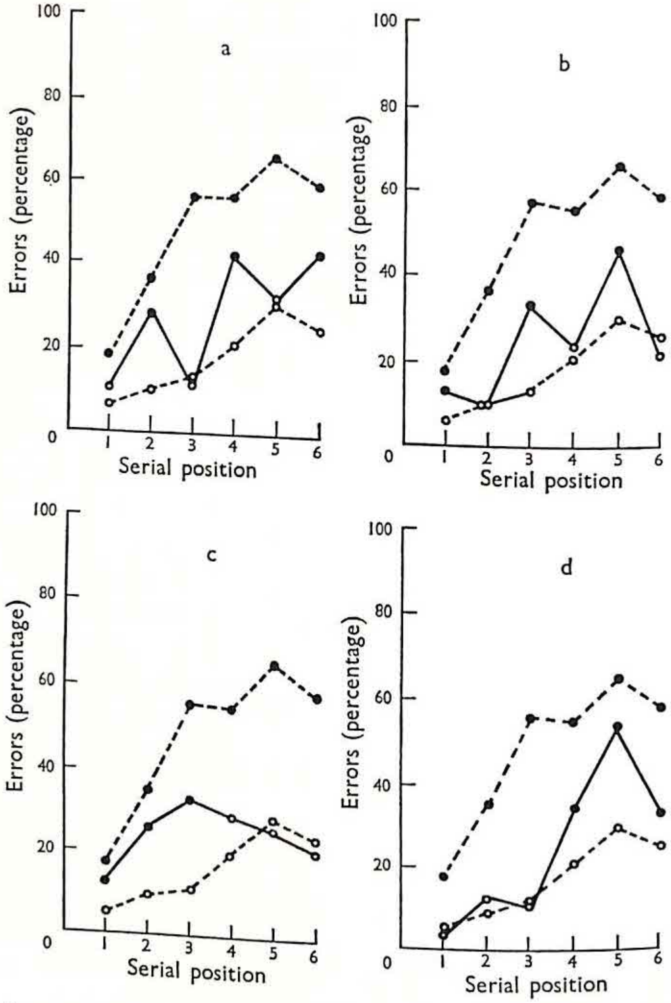
The subjects were 24 housewives and six naval ratings. They were tested in groups.

Results

Performance was scored in terms of per cent. errors, an error being any failure to reproduce the correct letter in the appropriate serial position. Mean per cent. overall error scores for the six conditions were as follows: A = 28·3; B = 25·0; C = 25·7; D = 25·8; E = 18·1; F = 49·2. There is a clear effect of acoustic similarity in E and F, the pure sequences, with 29 of the 30 subjects making more errors on F than on E. The Duncan Multiple Range Test showed no reliable differences among the four mixed conditions.

The main point at issue, however, is the distribution of errors rather than their overall frequency. This is shown in Figure 4; conditions A, B, C and D are shown in Figures 4*a*, 4*b*, 4*c* and 4*d* respectively. The pure sequence conditions E and F are included in each to provide reference points. The crucial question is whether errors

FIGURE 4



Mean per cent. errors as a function of serial position. Filled circles represent acoustically similar letters, open circles dissimilar letters. Conditions A, B, C and D are shown in Figs. 4*a*, 4*b*, 4*c* and 4*d* respectively. All figs. show conditions E and F as reference points.

occur predominantly *on* acoustically similar letters (filled circles) or on items immediately *following* similar letters. There is a marked tendency for errors to coincide with similar items, as the address hypothesis would predict. This shows up most clearly in the alternating conditions, A and B. The statistical reliability of the effect can be tested as follows. Each subject's score on letters occurring in positions 1, 3 and 5 is subtracted from his score on letters in positions 2, 4 and 6. On Wickelgren's hypothesis, this difference score should tend to be negative for condition A, and positive for condition B, whereas the address hypothesis would predict the opposite. Of the subjects, one shows a negative score on A, 25 show positive scores, and four show no difference, while on B, 21 show a negative score, four a positive score and five no difference, in other words the evidence overwhelmingly supports the address hypothesis rather than the associative hypothesis described ($\chi^2 = 17{\cdot}1$, $d.f. = 2$, $p < 0{\cdot}001$).

The serial position curves for conditions C and D also support this conclusion. The crucial positions here are 3 and 4, at the boundary between the similar and dissimilar blocks of letters. In condition D (DDDSSS), errors clearly increase from position 3 to position 4, whereas in condition C (SSSDDD), the opposite tends to occur. The reliability of this difference was assessed using as a score each subject's total errors on position 4 minus his errors on position 3. A Wilcoxon Test showed this to be significantly greater for condition D ($T = 60{\cdot}5$, $N = 25$, $p < 0{\cdot}01$, two-tail).

A striking feature of Figure 4 is the fact that for all mixed conditions, performance on acoustically different items is unaffected by the presence of acoustically similar items. The similar items, however, appear to be easier when mixed with dissimilar letters. The reason for this asymmetrical effect is not clear.

Experiment V gives clear support to an address hypothesis of retrieval, rather than to Wickelgren's associative hypothesis. Experiment VI aims to explore the generality of this result by using sequences of words rather than letters, and auditory rather than visual presentation.

EXPERIMENT VI

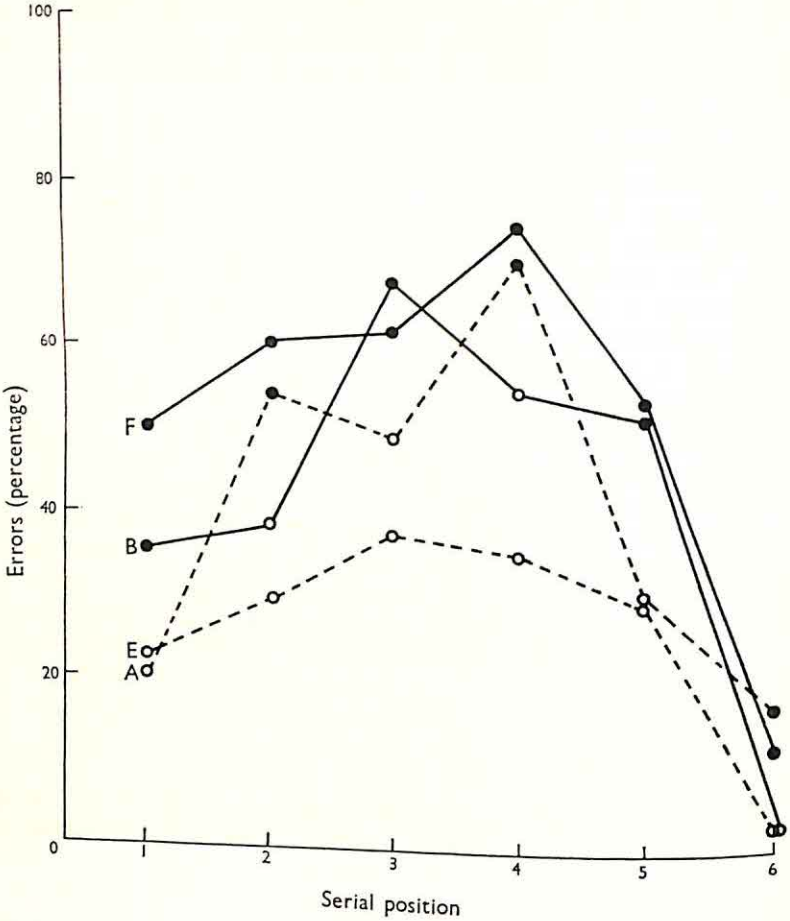
Method. Sequences of six words were made up from the two basic sets used in Experiment I. They comprised eight sequences of each of four types, A (DSDSDS), B (SDSDSD), E (DDDDDD) and F (SSSSSS). Words from the appropriate category were selected at random, with the constraint that a word could not occur more than once in any sequence. The 32 sequences were made into eight trial blocks, with each block containing the four types of sequence in random order.

The sequences were read out at a rate of 1 sec. per word; subjects were allowed 24 sec. per sequence to write their responses on a prepared sheet, followed by a warning signal, and 5 sec. later, the next sequence. Subjects were encouraged to guess when in doubt, and to maximize response availability the two sets of words were written on cards which were visible throughout the session. The subjects were 24 housewives from the A.P.R.U. panel and were tested in a single group.

Results

Scoring was again in terms of errors and the results are shown in Figure 5. Again there is a large overall effect of acoustic similarity, with all subjects, making more errors in condition F than in condition E. The within sequence effects also agree with Experiment V in that errors occur predominantly on similar items. This was again assessed by comparing scores on items, 1, 3 and 5 with scores on 2, 4 and 6. For condition A (DSDSDS), one subject made more errors on 1, 3 and 5, five subjects made the same number, and 18 made more errors on items 2, 4 and 6. For condition B (SDSDSD), the reverse occurred, with 23 subjects making more errors on items

FIGURE 5



Mean per cent. errors as a function of serial position for conditions A, B, E and F. Filled circles represent acoustically similar words, open circles acoustically dissimilar words.

1, 3 and 5 and one subject showing no difference. Again the results overwhelmingly support an address hypothesis rather than Wickelgren's associative hypothesis $(\chi^2 = 40{\cdot}8,\ d.f. = 2,\ p < 0{\cdot}001)$.

The overall picture presented by Experiment VI differs from that in Experiment V in two respects, however. When items are mixed, Figure 5 shows that similar items are of approximately the same level of difficulty as when they are presented in pure lists (condition F), while dissimilar words are apparently made more difficult. This is exactly the opposite to what occurred in the previous experiment. It is unclear whether this is due to using words rather than letters, a different mode of presentation, or larger sets of items (with sets of six and repetition excluded, pure lists in Experiment V always contained every item). The second striking difference is the much greater recency effect in Experiment VI, where the last item is consistently the easiest in contrast to Experiment V where the early items are easiest. This greater recency effect is characteristic of auditory presentation (Murray, 1966; Murdock, 1967), though in this case it may also be due to the fact that subjects in Experiment VI were not specifically instructed to recall in the order of presentation as they were in Experiment V.

Discussion

Both experiments show a marked tendency for errors in mixed lists to occur *on* rather than after acoustically similar items, a result which is clearly incompatible with the phonemic-associative hypothesis proposed by Wickelgren (1965*b*). This is perhaps not surprising when one considers the relative magnitudes of the acoustic similarity effect and the repetition effect on which the explanation is based. Acoustic similarity consistently produces a massive effect whereas the repetition effect which reflects the associative hypothesis in its purest and hence presumably strongest form is a relatively small and variable determinant of overall level of performance (Wickelgren, 1965*c*). The strongest evidence for the associative hypothesis comes from intrusions (Wickelgren, 1966) in a study which does not cite overall level of performance. Interpretation of intrusions must always be somewhat equivocal since they may simply represent an attempt by the subject to fill in gaps in his recall and may thus reflect guessing habits which may have little connection with the recall process (cf. Conrad, 1960).

How then *does* acoustic similarity influence short-term serial recall? We have so far considered seven hypotheses, three in terms of the input process, two in terms of storage and two retrieval hypotheses. Of these, only one is consistent with the data so far. Assuming that the various experiments have been adequately performed, and that they are fair tests of the various hypotheses, we can rule out six possibilities. This does not of course mean that the remaining explanation is therefore correct. It simply means that it has not yet been proved incorrect. In its present form the hypothesis is certainly too vague to be satisfactory, though it is capable of expansion so as to produce a number of more detailed sub-hypotheses, which in turn should produce a model of increasing specificity.

The author is grateful to Mrs. A. J. Hull for carrying out and analysing the results of Experiment V, and to Dr. N. S. Anderson and Dr. R. Conrad for useful discussion.

References

BADDELEY, A. D. (1966*a*). Short-term memory for word sequences as a function of acoustic, semantic and formal similarity. *Quart. J. exp. Psychol.*, **18**, 362–5.

BADDELEY, A. D. (1966*b*). The influence of acoustic and semantic similarity on long-term memory for word sequences. *Quart. J. exp. Psychol.*, **18**, 302–9.

BADDELEY, A. D., and DALE, H. C. A. (1966). The effect of semantic similarity on retroactive interference in long- and short-term memory. *J. verb. Learn. verb. Behav.*, **5**, 217–20.

BROADBENT, D. E. (1958). *Perception and Communication*. London: Pergamon Press.

BROWN, J. (1958). Some tests of the decay theory of immediate memory. *Quart. J. exp. Psychol.*, **10**, 12–21.

CONRAD, R. (1960). Serial order intrusions in immediate memory. *Brit. J. Psychol.*, **51**, 45–48.

CONRAD, R. (1963). Acoustic confusions and memory span for words. *Nature*, **197**, 1029–30.

CONRAD, R. (1964). Acoustic confusion and immediate memory. *Brit. J. Psychol.*, **55**, 75–84.

CONRAD, R., BADDELEY, A. D., and HULL, A. J. (1966). Rate of presentation and the acoustic similarity effect in short-term memory. *Psychon. Sci.*, **5**, 233–4.

CONRAD, R., and HULL, A. J. (1964). Information, acoustic confusion and memory span. *Brit. J. Psychol.*, **55**, 429–32.

CROSSMAN, E. R. F. W. (1959). Information and order in immediate memory. In *Information Theory*. London: Butterworth.

HINTZMAN, D. L. (1965). Classification and aural coding in short-term memory. *Psychon. Sci.*, **3**, 161–2.

LAUGHERY, K. R., and PINKUS, A. L. (1966). Short-term memory: Effects of acoustic similarity, presentation rate and presentation mode. *Psychon. Sci.*, **6**, 285–6.

MACKWORTH, J. F. (1964). Auditory short-term memory. *Canad. J. Psychol.*, **18**, 292–303.

MELTON, A. W. (1963). Implication of short-term memory for a general theory of memory. *J. verb. Learn. verb. Behav.*, **2**, 1–21.

MURDOCK, B. B. Jnr. (1965). Effects of a subsidiary task on short-term memory. *Brit. J. Psychol.*, **56**, 413–21.

MURDOCK, B. B. Jnr. (1966). Visual and auditory stores in short-term memory. *Quart. J. exp. Psychol.*, **18**, 206–11.

MURDOCK, B. B. Jnr. (1967). The effects of noise and delayed auditory feedback on short-term memory. *J. verb. Learn. verb. Behav.*, **6**, 737–44.

MURRAY, D. J. (1966). Vocalization at presentation and immediate recall, with varying recall methods. *Quart. J. exp. Psychol.*, **18**, 9–18.

PETERSON, L. R., and PETERSON, M. J. (1959). Short-term retention of individual verbal items. *J. exp. Psychol.*, **58**, 193–8.

POSNER, M. I., and ROSSMAN, E. (1965). Effect of size and location of informational transform upon short-term retention. *J. exp. Psychol.*, **70**, 496–505.

POULTON, E. C., and FREEMAN, P. R. (1966). Unwanted asymmetrical transfer effects with balanced experimental designs. *Psychol. Bull.*, **66**, 1–8.

RABBITT, P. M. A. (1966). Recognition memory for words correctly heard in noise. *Psychon. Sci.*, **6**, 383–4.

UNDERWOOD, B. J. (1949). *Experimental Psychology*. New York: Appleton Century Crofts.

WICKELGREN, W. A. (1965*a*). Similarity and intrusions in short-term memory for consonant-vowel digrams. *Quart. J. exp. Psychol.*, **17**, 241–7.

WICKELGREN, W. A. (1965*b*). Short-term memory for phonemically similar lists. *Amer. J. Psychol.*, **78**, 567–74.

WICKELGREN, W. A. (1965*c*). Short-term memory for repeated and non-repeated items. *Quart. J. exp. Psychol.*, **17**, 14–25.

WICKELGREN, W. A. (1966). Associative intrusions in short-term recall. *J. exp. Psychol.*, **72**, 853–8.

Manuscript received 16th April, 1968.

SENTENCE RETENTION AND TRANSFORMATION THEORY

BY

PATRICIA WRIGHT

From M.R.C. Applied Psychology Research Unit, Cambridge, England

The number of words recalled in addition to a sentence can be used to measure the amount of space a sentence occupies in immediate memory. Generally an active sentence (K) requires less space in memory than a passive negative sentence (PN). But in the present experiments it was found that when subjects were presented with a block of sentences of only one syntactic form, there was no longer a difference in the amount of space needed to store K and PN sentences. However, the difference between K and PN reappeared when the tense of the verb in successive sentences was varied, even though the material was still presented in a block of either all K or all PN sentences.

These results were interpreted as suggesting that subjects were remembering words and phrases from each sentence rather than deriving kernal strings and remembering transformation tags.

Introduction

There are a number of studies supporting the behavioural relevance of the linguistic notion of transformational generative grammar. Sentences that are more complex on grammatical measures are found more difficult for people to handle, i.e. to understand (McMahon, 1963) or remember (Mehler, 1963). There is also evidence for non-syntactic factors affecting the evaluation and retention of sentences. Wason (1965) has shown that the way in which subjects encode a stimulus display affects the time it takes to assess the truth or falsity of a statement about the display. Slobin (1966) has shown that children find passive sentences much more difficult if they are reversible, i.e. if it is semantically possible for the object of the verb to also be the subject.

Therefore there exists the possibility that effects attributed to syntax are due to psychological factors which may not be primarily linguistic at all. For example it is known that familiarity affects ease of retention (Dale, 1967). It could follow from this that if people are more familiar with some types of sentence, such sentences will be remembered more easily.

Strong support for the notion that sentences are remembered in accordance with the predictions of transformation theory comes from an experiment by Savin and Perchonock (1965). They measured the number of words which subjects could store in immediate memory while also remembering a sentence. For syntactically more complex sentences subjects could remember fewer words. Savin and Perchonock argued that sentences equivalent on transformation theory measures were equivalent in terms of the behavioural measures that they employed. The present series of experiments will examine whether processes other than those derived from transformation theory might give rise to similar effects for sentence retention.

In order to simplify the argument only two syntactic forms will be considered, the simple active indicative (K) and the passive transformation with the negation "not" inserted between the auxiliary and the main verb. The passive negative (PN) was chosen because it is a particularly difficult construction, involving two transformations, and required the most memory space of the sentence forms considered by Savin and Perchonock. The advantage of using such a difficult construction lies in the fact that there are large differences between performance on such sentences and performance on active sentences. If a certain amount of this difference is lessened, for example as a result of practice, there should still be an appreciable difference remaining that is the effect of the two transformations.

The following series of experiments first establishes that the material and procedure used give differences between K and PN, then various modifications of the procedure are examined.

General Method

Subjects heard a sentence, followed by eight words, and were required to write down the sentence and as many words as could be remembered. 30 sec. after the last word was given, subjects were told to turn over the page and 5 sec. later the next sentence was given. There were 21 tape-recorded sentences such as: The guard stopped the train. The meal was not prepared by the cook. The man believed the doctor. The money was not given by the government.

In all conditions and in all experiments the first three sentences were PN. The mean number of words recalled on these first three sentences was intended to provide a covariate or a basis for matching subjects, in order to increase the sensitivity of the statistical design when comparison was made between separate groups. The syntactic form of these three sentences was chosen quite arbitrarily. All subjects were told that the first three sentences would be PN and also the form of the remaining 18 sentences.

Although this technique was essentially that used by Savin and Perchonock (1965), there were several procedural differences. The same words and categories were used throughout this series of experiments but they were not identical to those used by Savin and Perchonock. The categories were altered because a pilot study using Savin's word lists indicated that many errors were intrusions from words given in earlier trials. Hoping to lessen this, categories were used from which 21 different words could be presented, so helping subjects to distinguish new from old items. Those of Savin's categories which could not be expanded to give 21 words were replaced. The categories used were: nature, furniture, number, body, name, weather, clothing, colour. An example of one such word list is "bush, lamp, ten, shoulder, Bob, rain, blouse, green."

The other major difference in procedure concerns the time interval between the sentence and the words. Savin and Perchonock left an interval of 5 sec. However, there were theoretical reasons for thinking that reducing this time interval would make the difference between K and PN greater. Therefore in the experiments reported below a gap of only 1 sec. was left between the presentation of the sentence and the lists of words. The words were presented at the rate of 1 per sec.

General procedure

Subjects were tested in groups of 15–22. They were instructed that they would hear a sentence followed by a list of eight words, each word being from a category in the list in front of them, and in the same order as the printed categories. They were told to wait until they had heard all the words, the last word always being a colour, and then to write down the sentence followed by as many words as they could still remember. Stress was laid on the importance of the sentence being correct, although words from the list could be written in any order and it did not matter if the words were written against the wrong category names.

Subjects

Men and women volunteers from the subject panel of the A.P.R.U. were used in all experiments. Their ages ranged from 20 to 69, with the majority being in the 30–50 year range.

Materials

Each subject was given a booklet consisting of 22 duplicated sheets. At the top of each sheet was the word "SENTENCE" followed by three blank lines. Below this was a side heading "WORDS" and a vertical list of the eight categories used. To the right of each category name was a blank line.

Experiment IA

Procedure

Following the first three PN sentences two groups of subjects were presented with a scrambled order of K and PN sentences. A sentence that was K for one group of subjects would be PN for another, and similarly the PN sentences were reversed. This was done to ensure that the differences in number of words recalled could not be due to variations in the content of sentences. Forty subjects took part in the experiment, 18 in one group and 22 in the other.

Results and discussions

The mean number of words remembered following K sentences was 3·62 (*s.d.* = 0·87) and following PN sentences was 3·21 (*s.d.* = 1·01). A sign test carried out on the means of K and PN for each subject indicated that this difference was reliable beyond the 1 per cent. level on a two-tailed test ($N = 40$, $x = 8$, $z = 3{\cdot}64$).

These differences are in the same direction as those obtained by Savin and Perchonock (1965). Therefore it might seem justified to assume that the materials and procedure being employed here are not substantially different from those used by Savin and Perchonock. However, the actual mean values obtained in the two experiments are somewhat different, particularly for the active sentences. Savin and Perchonock report a mean of 5·27 for K and 3·48 for PN. There may be many reasons why the figures obtained by Savin and Perchonock are generally higher than those of the present experiment (different sentences, different words, different subjects), but it seems more difficult to account for the much greater discrepancy with respect to the active sentences. Perhaps having the first three sentences PN created a "set" which, in spite of instructions about the form of the remaining sentences, made the processing of K sentences relatively more difficult. But such a factor might be expected to make the processing of PN easier and there is no evidence for this.

In a later experiment, Savin (1966) has argued that subjects require time to process a sentence and that if the word list is given immediately after the sentence then the observed decrement is due to both processing factors and storage factors. If this is so, one might expect that the differences between K and PN would be enhanced as a result of shortening the interval between the sentence and the list of words because on a transformation theory PN sentences require more complex processing than K sentences.

Experiment IB

In order to check on the effect of having a gap between the sentence and the word list, Experiment IA was repeated with a 5 sec. gap between the sentence and the words.

A further 34 subjects took part in this experiment.

Results

The mean number of words remembered following K sentences was 3·94 (*s.d.* = 1·13) and following PN sentences was 3·52 (*s.d.* = 0·98). Again a sign test carried out on the mean K and PN scores for each subject indicated that the difference between the two sentence forms was statistically reliable beyond the 1 per cent. level on a two-tailed test ($N = 34$, $x = 7$, $z = 3{\cdot}26$).

Comparing these results with the data obtained from Experiment IA bears out Savin's suggestion that the gap between the sentence and the word list increases the number of words recalled but, contrary to expectation, there was a similar improvement for the two sentence forms; for K sentences an improvement of 0·32 words and for PN sentences an improvement of 0·31 words.

There are at least two possible explanations of the failure to find differential improvement as a result of the delay between the sentence and the words. Either the "processing time" is so small a part of the 5 sec. interval that it has no measurable effect, or else subjects, when asked to remember and reproduce a sentence verbatim, do not process the sentence in accordance with the predictions derived from transformation theory.

Experimental evidence bearing on the first of these alternatives is equivocal. Miller and McKean (1964) obtained a processing time of 1·34 sec. when subjects were required to transform PN sentences into their equivalent actives, but Slobin found a difference of only 0·82 sec. in the harmonic means of the times to assess the truth of K or PN sentences. Therefore it is possible that the 1 sec. gap used above may have been sufficient to allow full processing of each sentence, in which case no additional effect of prolonging this interval between the sentence and the words would be expected. However, Gough (1966) reports data in support of the second of the alternative explanations being suggested. He found no effect of having either a zero or 3 sec. time interval between a sentence and the presentation of a picture either confirming or disconfirming the sentence. On the basis of his results Gough concluded that complex sentences were not reduced to their underlying structure the moment they were heard.

Although it is not possible to choose decisively between the alternative explanations of the present results, the main purpose of Experiment IB was to see if the procedure adopted in IA (i.e. having the words follow the sentence directly) lessened the observed differences between K and PN. Since it did not and since there were theoretical grounds for thinking that the procedure might enhance the effects of syntax it was decided to continue using the 1 sec. interval between the sentence and the word list.

Experiment II

If difficulty in handling PN sentences is primarily the result of lack of familiarity with such constructions, this might show in a greater improvement with practice for PN than is obtained from practising K sentences. Salzinger and Eckerman (1967) have data suggesting that with practice the differences between simple declarative sentences and passive negative questions completely disappear.

Procedure

Following the first three PN sentences subjects were presented with 18 sentences and following a short break another 21 sentences were given. Either the sentences were all PN, or, after the first three PN, all the remaining sentences were K. This change from the earlier procedure of scrambling the order of K and PN was made so that in one test session each subject could have more practice with a given syntactic form. Thirty-six subjects were presented with K sentences and 35 had PN sentences.

Results and discussion

The mean number of words recalled following the first 18 K sentences was 3·71 (*s.d.* = 0·31) and for the PN sentences was 3·84 (*s.d.* = 0·36). Although it had been intended to match subjects on the basis of their performance on the first three sentences it was found more satisfactory to make the comparison between the groups with reference to specific sentence contents. Comparing the two groups on the basis of the mean scores for each sentence, a sign test indicated no reliable differences between the groups: on seven sentences the K group had higher scores and on 11 the PN group were highest. The more sensitive Wilcoxon Signed Ranks test also showed no differences between the groups (obtained $T = 53$; for significance at the 5 per cent. level $T \leqslant 40$). Furthermore the difference in mean group scores contra-indicates a tendency for more words to be remembered after K than after PN sentences. Therefore it seems necessary to conclude that when subjects are presented with sentences of only one syntactic form there is no difference in the storage space required by K and PN sentences.

Although the experiment was intended to examine practice effects, this is clearly difficult when there were no differences between K and PN at the start of the experiment. Therefore the result obtained from the second block of 21 sentences is not being reported in detail, but there were still no significant differences between K and PN at the end of the second block of 21 trials.

That the differences between K and PN should disappear when subjects are presented with sentences of only one type is capable of several interpretations. Transformation theorists could argue that the experimental situation serves as the transformation store, since the subject knows which transformation will apply to all the sentences he will hear. That is to say, if subjects do not need to use storage space in immediate memory to retain the transformation, then K and PN would be expected to occupy the same amount of space.

Another interpretation, derived from the notion that PN sentences are difficult because they are unfamiliar, might be that as a result of a more frequent encounter with K sentences, subjects develop an expectancy for such sentences. This implies that subjects are in fact as capable of handling PN as they are K, but they generally start off on the wrong foot by expecting K. On this assumption, scrambling the order of K and PN as in Experiment I capitalizes on the expectancy effect, whereas presenting sentences of only one type eliminates it.

A third possibility is that subjects remember specific words in the sentence. Consider for example the PN sentence frame "the . . . was not. . . by the. . . ." When hearing sentences of only one syntactic form this sentence frame is redundant. It is only necessary to store in immediate memory the words filling the blanks. Since there will be three words for both K and PN sentences there should be no difference in the storage space required in either case.

It was found possible to distinguish between the effects of the redundancy of the transform and the redundancy of the sentence frame in a further experiment.

EXPERIMENT III

If an "expectancy" explanation is correct, or if subjects were using the experimental situation as the transformation store, then introducing variation in the word ordering of successive sentences while still presenting subjects with blocks of sentences of only one syntactic form should give the same results as Experiment II. On the other hand, if subjects had been capitalizing on the redundancy of the sentence frame, this would be impossible once variation in word ordering was introduced; in which case the difference in number of words remembered after K and PN sentences might again be observed.

Procedure

Experiment II was repeated with two modifications.

(*a*) Each subject heard only 21 sentences, instead of 42 as in Experiment II; but as in Experiment II, following three PN sentences the remaining 18 sentences were either all K or all PN.

(*b*) The tense and mood of the verb in successive sentences was varied, i.e. "was not . . . by," "will not be . . . by," "had not been . . . by," "might not have been . . . by," but across the two conditions the form of the verb in each sentence was the same, i.e. the tense of sentence 7 was future in both K and PN (e.g. The family will give the party. The party will not be given by the family), whereas the verb in sentence 8 was in the simple past tense in both K and PN (e.g. The village feared the chief. The chief was not feared by the village).

Comparison of the two PN sentences just given shows how varying the form of the verb disrupts the sentence frame "The . . . was not . . . by the . . ." Not only is the verbal auxiliary itself varying but also the location of "not" is now indeterminate; sometimes it immediately precedes the main verb, sometimes not.

Results and discussion

The mean number of words remembered following K sentences was 3·51 (*s.d.* = 0·36) and following PN sentences was 3·12 (*s.d.* = 0·29), a difference statistically reliable beyond the 1 per cent. level on a two-tailed sign test ($N = 18$, $x = 2$, $z = 3{\cdot}07$) when comparison is made of the performance on each sentence in the two conditions.

These results would seem to suggest that subjects were remembering specific words in Experiment II, since variation in the tense of the verb would not prevent the experimental situation still serving to store the syntactic transformation tag, if this is what subjects had previously been doing.

General Discussion

It can be argued that it is unreasonable to expect the theory of transformational generative grammar to account for data on sentence retention. Chomsky has made a clear distinction between competence and performance models (Chomsky, 1957), between what a person is capable of and what he does, and transformation theory is intended to be a model of competence, not of performance.

Nevertheless, as was mentioned in the Introduction, there have been instances where strikingly accurate predictions about performance have been made directly from the competence model. For example Savin and Perchonock (1965) made the strong prediction that sentences of comparable complexity on transformation measures are equivalent on behavioural measures. Savin and Perchonock report that independent estimates of the space required for say the passive (P) transformation, which can be obtained from noting the difference between K and P, N and PN, etc., were identical. They also report that estimates of the space required to store one transformation were comparable regardless of whether the transformation was from N or K or from P to K. Such data would seem to suggest that very little modification of the competence model is necessary to accurately predict features of performance. But the present data fits less easily into the picture.

If the similarity of K and PN, when these are presented in blocks of syntactically similar sentences, is to be accounted for within the framework of transformation theory (i.e. with the input being stored as a kernal string and a transformation tag) it seems necessary to suggest that subjects might be using the experimental situation itself to store the transformation tag; but it then becomes unclear why this is no longer done when the tenses of the verbs are varied. Miller and McKean (1964) have shown that expanding the verbal auxiliary increases slightly the time required to make specified transformations of a sentence, but they did not examine the interaction of this expansion with the voice of the sentence.

Of course it is arguable that the results of Experiment II should be discounted on the grounds that if subjects were remembering particular words then they were not responding to the sentence as a sentence; in which case transformation theory could not be expected to apply. The main difficulty that such argument encounters is that once it has been admitted that some sentences can be remembered and reproduced on the basis of *words* rather than larger linguistic units, how can one be certain that this is not true of all sentences.

Nevertheless, the point is well taken that the experimental situation in Experiment II may have been such that subjects had available, and made use of, strategies of memorization that are not typical of sentence retention in general. Perhaps the only defence against this criticism is to compare performance across the three experiments. For the K sentences there was little difference in the mean number of words remembered in Experiments IA, II and III; the mean scores were 3·62, 3·71, 3·51 respectively. Whereas for the PN sentences the differences were much greater between the experiments (3·21, 3·84 and 3·12). The similarity of performance on the K sentences might be interpreted as indicating that subjects were performing in an essentially similar manner throughout these experiments. Accepting this interpretation means that the results of Experiment II must be accounted for in terms that are generalizable over all three experiments.

An explanation in terms of subjects remembering words or phrases might imply that subjects were able to capitalize on the greater redundancy of the PN sentences in the procedure of Experiment II than in the procedure adopted in the other experiments. For the K sentences the redundancy concerns only the two definite articles, and this sentence frame was in fact the same throughout the three experiments. For the PN sentences the additional redundancy concerns the phrase "was not . . . by." This phrase was redundant only in Experiment II. Therefore it might be the space occupied by this phrase that results in the available memory space generally being less after PN sentences.

An explanation in terms of subjects remembering specific phrases might not account for the similarities that Savin and Perchonock observed in the effect of different transformations. Nevertheless, further support for the suggestion that subjects were not "transforming" the original PN sentence comes from the data obtained in Experiment IB, where it was not possible to detect a difference between the two sentence forms as a function of introducing a delay between the sentence and the word list. If the sentences were undergoing processing of different complexities it would seem reasonable to expect that the delay would result in a greater improvement in the number of words recalled after PN sentences (the more difficult to process). No differential improvement was found, and this implies no differential processing of K and PN sentences.

Therefore, taking this series of experiments as a whole, it would seem that the competence model of transformational generative grammar did not always lead to accurate predictions of performance. Nor was it evident that any modification of such a model would suffice to account for the present data. This raises the question of whether it might be easier to encompass sentence retention within a psychological theory of retention rather than within a primarily linguistic theory. For example the retention of phrases might correspond to the "*adopted chunks*" of Tulving and Patkau (1962). Already McNulty (1966) has shown how usefully the notion of adopted chunks can be applied to the retention of verbal material of varying approximation to English, and Epstein (1967) has shown that when subjects are trying to learn syntactically structured but anomolous sentences, cues that help them to "chunk" the material increase the number of words in the sentence that can be recalled. The increase was greater for syntactically more complex sentences, passives and nominalizations, than for simple active sentences; a finding that ties in with the result of Experiment II above, where marked improvement in performance with the difficult PN sentences was noted.

Epstein obtained his "chunks" by segmenting the material at phrase junctures, so his results could be taken as support for the notion that syntax affects sentence retention by altering the phrase structure of the sentence, which in turn determines into how many "chunks" the sentence is broken. The general relationship between retention and both number of "chunks" and the number of "items per chunk" has been established by Mandler (1967). It would require further experimentation to establish whether or not, so far as sentence retention is concerned, simple sentences can validly be considered as perhaps three "chunks" (NP–VP–NP), with the number of "items per chunk" for K sentences being respectively 2–1–2 (The cat–saw–the dog) and for PN sentences being 2–3–3 (The dog–was not seen–by the cat). Practical difficulties may prevent any straightforward application of the "chunking" model, for example it may be quite untenable to assume uniform difficulty among the items within a "chunk." Indeed there is evidence (Roberts, 1967) that nouns are less subject to forgetting than adjectives or adverbs. Nevertheless the data from the present experiments supports the suggestion that sentence retention can be considered from the vantage point of the psychology of retention, and is perhaps more helpfully viewed in this light than from the standpoint of a linguistic theory of competence.

I am very grateful to Dr. A. D. Baddeley for suggesting the modification adopted in Experiment III.

REFERENCES

CHOMSKY, N. (1957). *Syntactic Structures.* The Hague: Mouton.

DALE H. C. A. (1967). Familiarity and free recall. *Quart. J. exp. Psychol.*, **19**, 103–8.

EPSTEIN, W. (1967). Some conditions of the influence of syntactical structure on learning: grammatical transformation, learning instructions, and "chunking." *J. verb. Learn. verb. Behav.*, **6**, 415–9.

GOUGH, P. B. (1966). The verification of sentences: The effects of delay of evidence and sentence length. *J. verb. Learn. verb. Behav.*, **5**, 492–6.

MANDLER, G. (1967). Organisation and memory. To appear in SPENCE, K. W. and SPENCE, J. T. (Eds.). *The Psychology of Learning and Motivation: Advances in Research and Theory.* New York: Academic Press.

MCMAHON, L. E. (1963). *Grammatical Analysis as Part of Understanding a Sentence.* Unpublished doctoral dissertation, Havard University.

MCNULTY, J. A. (1966). Measurement of "adopted chunks" in free recall learning. *Psychon. Sci.*, **4**, 71–2.

MEHLER, J. (1963). Some effects of grammatical transformation on the recall of English sentences. *J. verb. Learn. verb. Behav.*, **8**, 250–69.

MILLER, G. A., and MCKEAN, K. O. (1964). A chronometric study of some relations between sentences. *Quart. J. exp. Psychol.*, **16**, 297–308.

ROBERTS, K. (1967). Some effects of grammatical context on the recall of sentence elements. Paper presented at the Midwestern Psychol. Ass. Convention, Chicago, Illinois.

SALZINGER, K., and ECKERMAN, C. (1967). Grammar and the recall of chains of verbal responses. *J. verb. Learn. verb. Behav.*, **6**, 232–9.

SAVIN, H. B. (1966). Grammatical structure and the immediate recall of English sentences: II Embedded clauses. Mimeo, University of Pennsylvania.

SAVIN, H. B., and PERCHONOCK, E. (1965). Grammatical structure and the immediate recall of English sentences. *J. verb. Learn. verb. Behav.*, **4**. 348–53.

SLOBIN, D. I. (1966). Grammatical transformations and sentences comprehension in childhood and adulthood. *J. verb. Learn. verb. Behav.*, **5**, 219–27.

TULVING, E., and PATKAU, J. E. (1962). Concurrent effects of contextual constraint and word frequency on immediate recall and learning of verbal material. *Canad. J. Psychol.*, **16**, 83–95.

WASON, P. C. (1965). The contexts of plausible denial. *J. verb. Learn. verb. Behav.*, **4**, 7–11.

Manuscript received 25th March, 1968.

REASONING ABOUT A RULE

BY

P. C. WASON

From Psycholinguistics Research Unit, University College London

Two experiments were carried out to investigate the difficulty of making the contrapositive inference from conditional sentences of the form, "if P then Q." This inference, that not-P follows from not-Q, requires the transformation of the information presented in the conditional sentence. It is suggested that the difficulty is due to a mental set for expecting a relation of truth, correspondence, or match to hold between sentences and states of affairs. The elicitation of the inference was not facilitated by attempting to induce two kinds of therapy designed to break this set. It is argued that the subjects did not give evidence of having acquired the characteristics of Piaget's "formal operational thought."

INTRODUCTION

This investigation is concerned with the difficulty of making a particular type of inference from conditional sentences, statements of material implication of the form, "if P then Q" ($P \supset Q$).

Within the propositional calculus the truth table for the conditional counts the following combinations of the components of the sentences as true: PQ, $\bar{P}Q$, $\bar{P}\bar{Q}$ (where $\bar{P}$ stands for not-P and $\bar{Q}$ for not-Q), and only the one combination, $P\bar{Q}$, as false. It follows that only values of P and values of $\bar{Q}$ allow a valid inference. In other words a valid inference depends crucially upon the *possibility* of meeting the falsifying contingency, $P\bar{Q}$. It rests simply upon the denial of this contingency. $\bar{P}$ comes out true whether it is associated with Q or $\bar{Q}$,and Q comes out true whether it is associated with P or $\bar{P}$. For example, suppose someone says, "if John is a stockbroker, then he is bound to read the *Financial Times*." If John isn't a stockbroker, or if he does read the *Financial Times*, the conditional statement (to the logician) is inevitably true. But if he is a stockbroker, or if he never reads the *Financial Times*, then there are grounds for verifying or falsifying the conditional statement. Hence, logically, only two forms of inference are valid: "$P \supset Q$ and $P \ldots Q$" (*modus ponens*) and "$P \supset Q$ and $\bar{Q} \ldots \bar{P}$" (*modus tollens* or contrapositive). The other two forms of inference, "$P \supset Q$, and $\bar{P} \ldots \bar{Q}$" (denial of the antecedent) and "$P \supset Q$ and $Q \ldots P$" (affirmation of the consequent) are fallacious.

In a task consisting of sentences expressed in everday terms the author (Wason, 1964) has shown that the affirmation of the consequent occurs significantly more frequently than the denial of the antecedent. In an abstract task the occurence of all four types of inference has been investigated in a pilot study (Wason, 1966). The subjects were presented with the following sentence, "if there is a vowel on one side of the card, then there is an even number on the other side," together with four cards each of which had a letter on one side and a number on the other side. On the front of the first card appeared a vowel (P), on the front of the second a consonant ($\bar{P}$), on the front of the third an even number (Q), and on the front of the fourth an odd number ($\bar{Q}$). The task was to select all those cards, but only those cards, which would have to be turned over in order to discover whether the experimenter was lying in making the conditional sentence. The results of this study, and that of a replication by Hughes (1966), showed the same relative frequencies of cards selected. Nearly all subjects select P, from 60 to 75 per cent. select Q, only a minority select $\bar{Q}$ and hardly any select $\bar{P}$. Thus two errors are committed: the consequent is fallaciously affirmed and the contrapositive is withheld. This type of task will be called henceforth the "selection task." These errors seem to be enduring and deep. Hughes has shown that they often persist for 15 trials, even when the subjects turn over all the cards after each trial and evaluate the conditional sentence with respect to them.

A theory to explain these results has been postulated (Wason, 1966). It rests on two assumptions. The first assumption is that individuals are not constrained by the rules of the propositional calculus. They implicitly assume that a conditional sentence can have three outcomes or truth values. PQ is true, $P\bar{Q}$ is false and $\bar{P}$ with either Q or $\bar{Q}$ is irrelevant. This interpretation is not original: it has been debated in the history of logic since the Stoic and Megarian schools. The assumption explains why the consequent is affirmed—Q is selected in order to see whether it is associated with P making the conditional true. It also explains why the antecedent is so infrequently denied—$\bar{P}$ is irrelevant to the truth or falsity of the sentence.

The second assumption explains the infrequency of the contrapositive inference—why $\bar{P}$ is so seldom deduced from $\bar{Q}$. This assumption is that individuals are biased, through a long learning process, to expect a relation of truth, correspondence or match to hold between sentences and states of affairs. In adult experience truth is encountered more frequently than falsity, and we seldom use a proposition or judgement that something is false in order to make a deduction. The semantic concept of falsity is logically equivalent to the syntactic concept of negation, and it has been shown that both cause difficulty when sentences have to be evaluated or constructed (Wason, 1959, 1961; Wason and Jones, 1963). Both concepts are relevant to the selection task. A value of $\bar{Q}$ represents a mismatch between a state of affairs visible on the card and a clause in the conditional sentence. This mismatch has to be recognized as such by transforming the information in the relevant clause from Q to $\bar{Q}$. In other words, a state of affairs (x) has to be seen not simply as x but as $\bar{Q}$. In doing this the individual presumably makes a judgement of falsity by uttering a covert negative sentence to himself. The difficulties involved in doing all this are assumed to be sufficient to account for the relative failure to select $\bar{Q}$.

According to this theory the affirmation of the consequent—the deduction of P from Q—is a plausible inference. But the withholding of the contrapositive remains irrational and is consequently a factor of much greater interest. The present investigation is concerned with the effects of therapies designed to correct the bias towards truth or correspondence, and thus facilitate the elicitation of the contrapositive inference.

EXPERIMENT I

The projection of falsity

It was predicted that if individuals were to "project falsity," i.e. to say what values, if any, associated with the given values, P, $\bar{P}$, Q, $\bar{Q}$, would make a conditional sentence *false*, then they would more readily select $\bar{Q}$ to determine whether the sentence was in fact true or false. P. is the only value, associated with $\bar{Q}$, which could make a conditional sentence false. Hence if individuals were to project P on to a value of $\bar{Q}$, they might be more likely to select $\bar{Q}$ as informative.

Method

Design. In the experimental group the subjects first carried out a selection task. They decided which of four cards (P, $\bar{P}$, Q, $\bar{Q}$) would enable them to determine whether a given conditional sentence was true of false, if they were to know the values on the back of the cards. They were then invited to say what values, if any, on the back of the cards would render the sentence false. They then revised their initial selection of the cards, if they wished to do so, and finally evaluated the sentence with respect to the values on both sides of all four cards, i.e. they turned over each card and evaluated the sentence as true or false with respect to it. In the control group the procedure was similar, but instead of projecting falsity the subjects were simply asked to think again about their initial selections, i.e. to revise them, "because people often do this task too quickly and get it wrong."

Subjects. Thirty-six first year psychology and statistics students of University College London.

Material. Two conditional sentences were typed on separate cards (5 × 3 in.): (1) "If there is a D on one side of any card, then there is a 3 on its other side," (2) "If there is a 3 on one side of any card, then there is a D on its other side." These two sentences were used as a control for the order in which the two items were mentioned. Associated with each sentence were four cards (2 × 2 in.). The cards associated with sentence (1) had the following letters and numbers on their front, and (in brackets) the following on their back: D(3), 3(K), B(5), 7(D). The following cards were associated with sentence (2): 3(D), K(3), 5(B), D(7). It will be appreciated that both sets of cards conform to only one of the two possible combinations of values of the antecedent and consequent: $P(Q)$, $Q(\bar{P})$, $\bar{P}(\bar{Q})$, $\bar{Q}(P)$. The other combination was not used: $P(\bar{Q})$, $Q(P)$, $\bar{P}(Q)$, $\bar{Q}(\bar{P})$.

Procedure. The subjects were allocated alternately to the groups and tested individually. Half the subjects in each group were tested with sentence (1) and half with sentence (2). The sentence was placed on the desk and the four test cards were placed in a line in a random order face upwards in front of the sentence. The subjects were told that cards with letters on their front had numbers on their back and *vice versa*. In the selection task the experimenter pointed to each card in turn and asked the subject whether knowing what was on the other side would enable him to find out whether the sentence was true or false. During the projection of falsity, in the experimental group, the experimenter pointed to each card and asked the subject to name a letter (or number) on the other side which would make the sentence false, or to say "none," if none would do so.

RESULTS

Two subjects, both in the control group, seemed unable to comply with the instructions and were hence rejected.

Tables I and II show the frequency of the responses in the different phases of the task for the experimental and control groups respectively. It will be noted immediately that, by comparing the frequencies in the diagonal cells, there is a greater tendency for the initial and revised selections of the cards to conform in the control group than in the experimental group. Inspection shows that the frequency of selecting $\bar{Q}$ increases over the task from five to eight in the experimental group and from two to three in the control group. The prediction that the selection of $\bar{Q}$ would be facilitated significantly in the experimental group is not confirmed.

Table I shows that in the experimental group 13 subjects did not select $\bar{Q}$ initially, and that of these, five did not project P on to $\bar{Q}$ and eight did project it. But of these eight, only three included a value of $\bar{Q}$ in their revised selection. Thus the therapy of falsifying the values cannot always be induced, and even when it is induced it is by no means effective.

Table I shows that the most frequent projection of falsity was on to all four values (P, $\bar{P}$, Q, $\bar{Q}$). But only one of the six subjects responsible for doing this subsequently selected all four values in their revised selection. It is evident that in these cases falsification did not render all the values acceptable for testing the truth or falsity of the sentence. The invitation to project falsity in these cases seemed to result in an arbitrary or indiscriminate response which had no bearing on subsequent behaviour.

It will be noted that P and Q is selected with the greatest frequency in both the experimental group (eight cases) and in the control group (10 cases).

TABLE I

FREQUENCY OF INITIAL SELECTIONS, PROJECTIONS OF FALSITY AND REVISED SELECTIONS IN THE EXPERIMENTAL GROUP

Initial selection	*Projection of falsity*	Revised selection							
		PQ	P	PQQ̄	PP̄Q	PP̄QQ̄	PQ̄	N	N
P Q	P Q	2				1		3	8
	P P̄ Q Q̄	4				1		5	
P	P Q	1						1	4
	P	1						1	
	P Q̄						1	1	
	P P̄ Q Q̄	1						1	
P Q Q̄	P		1					1	3
	P Q̄			1				1	
	P Q Q̄			1				1	
P P̄ Q	P Q̄						1	1	1
P P̄ Q Q̄	P̄ Q̄						1	1	2
	P Q Q̄						1	1	
	N	9	1	2	0	2	4	18	18

TABLE II

FREQUENCY OF INITIAL AND REVISED SELECTIONS IN THE CONTROL GROUP

Initial selection	*Revised selection*						
	PQ	P	PQ̄	PP̄	PQQ̄	PP̄QQ̄	N
P Q	9					1	10
P	1	2					3
P Q̄			1				1
P P̄				1			1
P Q Q̄					1		1
P P̄ Q Q̄							0
N	10	2	1	1	1	1	16

Table III is the frequency distribution of the evaluation of the contingencies as true, false or irrelevant for the combined groups.

It will be noted that the author's theory is corroborated with respect to the evaluation of the PQ, PQ̄ and P̄Q̄ contingencies, but is refuted with respect to the evaluation of the P̄Q contingency. This was evaluated as making the sentence false by 22 subjects and as being irrelevant by only 10 subjects. This result is particularly

TABLE III

Frequency of the Evaluations in the Combined Groups

PQ	$\bar{P}Q$	$P\bar{Q}$	$\bar{P}\bar{Q}$	N
t	f	f	i	15
t	i	f	i	9
t	f	f	t	3
t	t	f	t	2
t	f	f	f	2
i	f	f	i	1
t	f	i	i	1
t	i	t	i	1
				34

t = true, f = false, i = irrelevant.

interesting because, on the revised selection, $\bar{P}$ was selected only four times out of 34, and yet, when it is associated with Q there is a much greater tendency for it to be judged as relevant to the falsity of the sentence.

This experiment has shown the inadequacy of projecting falsity as a therapy for the elicitation of the contrapositive inference, and that some subjects cannot even perform the therapeutic exercise. Two further pilot studies were carried out which revealed one interesting phenomenon. In the first study the four values were presented on separate trials and a few subjects, when presented with $\bar{P}$ and $\bar{Q}$, said they did not need to turn these over because they already falsified the conditional sentence. In the second study the subjects were asked to pick out only those values which "could break the rule" (i.e. falsify the conditional sentence). Four subjects selected *only* values of $\bar{P}$ and $\bar{Q}$ and refused to turn them over because they claimed this was useless. "It doesn't make any difference—the two I have chosen do break the rule." "There is no rule regarding that card ($\bar{P}$)." Thus in a small minority of subjects the concept of something following a rule appears to be inadequately conceived, for to know what could follow a rule is to know what could break that rule.

Experiment II

The restricted contingency programme

It was predicted that if subjects were initially allowed to evaluate examples of the four contingencies with respect to a given conditional sentence, and were, in addition, told that only one contingency falsified the sentence, then they would subsequently select values of $\bar{Q}$ within the same task to a greater extent than those who had not had this experience. The term, "restricted," is used because the intention of this programme was not to teach the truth table for the conditional, but to make the subject aware that only the $P\bar{Q}$ contingency falsified the sentence. It was reasoned that if the subject knows in advance that $\bar{Q}$ is crucial for falsification, then he might select it as potentially informative.

Method

Design. The experimental group first received the programme and then carried out a selection task with the same material. The control group carried out a selection task without receiving the programme.

Subjects. Twenty-six first year psychology students at the University of Edinburgh.

Material. Two conditional sentences were typed on separate cards similar to those used in Experiment I. (1) "If there is a square on one side of the card, then there is a red scribble on the other side," (2) "If there is a red scribble on one side of a card, then there is a square on the other side." Four programme cards were prepared for the experimental group which had the following stimuli on either side: (*a*) square, yellow scribble, (*b*) square, red scribble, (*c*) rectangle, red scribble, (*d*) hexagon, brown scribble. Eight similar cards were prepared for the selection task, the items mentioned first being on the front of the cards, and the items mentioned second (in brackets) being on the back: (*a*) square (red scribble), (*b*) square (brown scribble), (*c*) red scribble (square), (*d*) red scribble (hexagon), (*e*) green scribble (rectangle), (*f*) parellelogram (yellow scribble), (*g*) triangle (red scribble), (*h*) blue scribble (square). It will be appreciated that these cards represent both combinations of values of the antecedent and consequent: $P(Q)$, $P(\bar{Q})$, $\bar{P}(Q)$, $\bar{P}(\bar{Q})$, $Q(P)$, $Q(\bar{P})$, $\bar{Q}(P)$, $\bar{Q}(\bar{P})$.

Procedure. The subjects were allocated alternately to the groups and tested individually. Six subjects in the experimental group were tested with sentence (1) and seven with sentence (2), these proportions being reversed in the control group. Before presenting the conditional sentence the subjects in both groups examined the cards briefly to familiarize them with the fact that there was always a geometric shape on one side and a coloured scribble on the other side.

In the experimental group the conditional sentence was presented and the four programme cards were handed to the subjects who were asked to pick out "the one card which makes the rule false" (i.e. falsifies the conditional sentence). They were then asked to pick out any which "prove the rule true." It was explained to them that their decisions meant that the converse of the sentence could not be assumed—"that the rule only held one way." The subjects in the control group were given a similar amount of time to understand the conditional sentence without any explanation of its meaning.

The selection task had three phases which were the same for both groups.

Phase 1. The eight selection task cards were placed on the desk front side up in a random array, and the subject was asked to pick out "all those cards, but only those cards, which would show you, if you knew what was on the other side, that the rule was true or false."

Phase 2. The subject was asked to turn over all those cards which he had selected and evaluate the conditional sentence with respect to each: "tell me whether each proves the rule true or proves it false."

Phase 3. The subject was invited to project falsity on to the residual cards i.e. those cards which had not been selected. Each pair of cards (representing the same value) was pointed to in turn, starting with the $\bar{Q}$ cards, if these had not been selected: "could anything on the back of those cards make the rule false?"

RESULTS

In the programming in the experimental group all the subjects picked out the $P\bar{Q}$ card without hesitation as the only falsifying instance, and they all picked out PQ as the only verifying instance.

Table IV is the frequency distribution of the choices made in the selection task for both groups. (All subjects were consistent in picking out both instances of any value which they selected).

It is at once apparent that there is little difference in the results of the two groups and that the treatment given to the experimental group failed in its effects.

Table V is the frequency distribution of the evaluation of the contingencies and the projections of falsity on to the residual values in the combined groups. An empty cell in the evaluation task means that at least one of the values associated with that

TABLE IV

FREQUENCIES OF VALUES SELECTED IN BOTH GROUPS

	Experimental group	*Control group*	*N*
P Q	6	6	12
P	5	4	9
P Q̄	1	1	2
P P̄	1	1	2
P P̄ Q Q̄	0	1	1
			26

TABLE V

FREQUENCY OF THE EVALUATIONS OF THE CONTINGENCIES AND OF THE PROJECTION OF FALSITY ON RESIDUAL VALUES IN THE COMBINED GROUPS

Evaluation task				*Falsification task*			
PQ	P̄Q	PQ̄	P̄Q̄	P̄	Q	Q̄	N
t		f		—	—	P	6
t	f	f		—		—	5
t	f	f		—		P	3
t	i	f		—		P	2
t		f	i	—	—		2
t	f	f		Q		P	2
t	f	f	t		P̄	P	1
t		f		—	—	—	1
t		f		—	P̄	P	1
t	i	f	i		—	P	1
t		f		Q	P̄	P	1
t	f	f	i				1
							26

t = true, f = false, i = irrelevant. Note that no P column is included under falsification because all subjects selected P.

cell was absent from both sides of a selected card. An empty cell in the falsification task, however, means that a value on the front of a card, associated with that cell, had been selected (and evaluated), and hence was no longer available for the projection of falsity. A dash in a cell means that a subject denied that any value, associated with the value in that cell, would falsify the rule, and an entry of a value in a cell means that it, associated with the value in that cell, would falsify the rule. It will be noted that the selected values, other than P which was always selected, correspond to the *empty* cells in the falsification task.

It is of particular interest to note that six out of a possible 17 subjects failed to say that P, projected on to $\bar{Q}$, would falsify the rule; and that eight of the 10 subjects, who had claimed that $Q(\bar{P})$ falsified the rule, subsequently said that no value associated with $\bar{P}$ would falsify the rule.

Only nine of the 24 subjects, who had not selected $\bar{Q}$, said that they were aware of having made a mistake in the selection task.

General Discussion

The results show that two kinds of therapy do not facilitate the act of making the contrapositive inference, the selection of $\bar{Q}$, but they do show phenomena of considerable interest. The selection task was not meaningless to the subjects. Their results are far from random. In the combined experiments 50 per cent. initially select just P and Q out of the possible 15 combinations of values which could have been selected. This marked tendency to pick out only those values which are mentioned in the conditional sentence suggests that the selection task seemed deceptively easy. Its real meaning, the challenge which it implies, escaped the subjects to a large extent. But the selection of P and Q, which is so resistant to therapy, is consistent with the theory that individuals are biased through a long learning process to seek and expect a simple correspondence to hold between sentences and states of affairs. The introspections corroborate the theory. "You first of all accept all the cards as true—you don't make any allowances for any of them being wrong until you turn them over." "A rule is a rule, so looking at it frankly the ones with squares will have red scribbles on the back." "I feel very unhappy about my original choice, but yes, I would still choose the same ones if I had to do the task again." One subject twice projected truth instead of falsity before he could be prevailed upon to comply with the instructions.

It is a reasonable inference that this set for truth inhibits the perception of a card as being an exemplar of $\bar{Q}$. Even when it is recognized as such it is not always used to make a deduction. In the results of the combined experiments 16·7 per cent. of the subjects select $\bar{Q}$. But of those who do not initially select it, 30·6 per cent. fail to project P on to it as a means of falsifying the conditional sentence. This seems extraordinarily capricious. It is as if someone had said that any number other than 3 on the other side of D would falsify the conditional and in the next breath denied that any letter on the other side of 7 would do so. However, once the drift of the subjects' reasoning is apparent the logical discrepancies begin to look more plausible. P(Q) and Q(P) come to the same thing when the cards are turned over: they both verify the rule. And when $Q(\bar{P})$ is turned over it certainly doesn't verify the rule—hence it must falsify it. But of course there is nothing on the other side of $\bar{P}$ which would falsify the rule because there is nothing on the other side which could verify it. One might risk a general statement about this. Suppose that a rule is confirmed whenever one state of affairs (y) depends upon another (x); and suppose initially one has access to either x or y but not both. If y is obtained, then there is an expectancy that x will be found to have been associated with it, thus confirming the rule; but if x is found not to have been associated with y, we are tempted to say that the rule has been refuted. However, if initially x is not obtained, no such expectancy is generated, and we are less inclined to say that, if y were to follow, the rule would be refuted. And if y is not obtained initially nothing seems to follow about the rule at all.

The results, however, are still disquieting. If Piaget is right (Inhelder and Piaget, 1958), then the subjects in the present investigation should have reached the stage of formal operations. A person who is thinking in these terms will take account of the possible and the hypothetical by formulating propositions about them. He will be able to isolate the variables in a problem and subject them to a combinatorial analysis. But this is exactly what the subjects in the present experiment singularly fail to do. The variables in the present tasks are abstract but they are distinct and susceptible to symbolic manipulation. Could it then be that the stage of formal operations is not completely achieved at adolescence, even among intelligent individuals?

It may not, however, be the concept of implication which causes difficulty so much as its customary verbal guise: "if P then Q." That form of words suggests that the consequent follows the antecedent in time, or even that there is a causal relation between them. Other expressions have been tested. Hughes (1966) has shown that the logically equivalent expression, "Q if P," causes, if anything, even more difficulty. But implication can be formulated with simpler logical connectives, e.g. "$\bar{P}$ or Q" (either not-P or Q), "-(P$\bar{Q}$)" (it is not the case that P and not-Q). P. N. Johnson-Laird (personal communication) has reported that the mode of formulation makes a considerable difference when the contingencies have to be evaluated. We intend to test these and other formulations to see whether similar differences can be detected in selection tasks. This problem may be generalized. Is it the formal structure of rules which is responsible for their difficulty, or is it the words with which we express these rules? And if the latter, what words illuminate the structure?

I am particularly indebted to my colleague, Dr. P. N. Johnson-Laird, for an invaluable criticism of earlier drafts of this paper. I should also like to thank Dr. Margaret Donaldson and Mr. Roger Wales, of the University of Edinburgh, for kindly organizing the subjects and generously giving me facilities for conducting Experiment II. Finally, I owe a debt to my subjects. Their interest and enthusiasm has been a source of constant gratification.

REFERENCES

HUGHES, M. A. M. (1966). *The Use of Negative Information in Concept Attainment.* Unpublished University of London Ph.D. thesis.

INHELDER, B., and PIAGET, J. (1958). *The Growth of Logical Thinking.* New York: Basic Books.

WASON, P.C. (1959). The processing of positive and negative information. *Quart. J. exp. Psychol.*, **11**, 92–107.

WASON, P. C. (1961). Response to affirmative and negative binary statements. *Brit. J. Psychol.*, **52**, 133–42.

WASON, P. C., and JONES, S. (1963). Negatives: denotation and connotation. *Brit. J. Psychol.*, **54**, 299–307.

WASON, P. C. (1964). The effect of self-contradiction on fallacious reasoning. *Quart. J. exp. Psychol.*, **16**, 30–4.

WASON, P. C. (1966). Reasoning. In Foss, B. (Ed.), *New Horizons in Psychology.* Harmondsworth: Penguin Books.

Manuscript received 2nd April, 1968.

HANDWRITING, TYPEWRITING AND HANDEDNESS

BY

K. A. PROVINS* and D. J. GLENCROSS

From Departments of Psychology and Physical Education, University of Adelaide

The performance of two groups of 20 trained touch typists and one group of 20 non-typists was tested on seven different typewriting exercises. Three of these exercises were executed entirely with the right hand and three were completed using only the left. A comparison of performances of the two sides showed, for the typists, no differences between the sides or a difference in favour of the left hand; and for the non-typists, a difference between hands in favour of the right side in two of the three comparisons.

A comparison of handwriting performance with the preferred and non-preferred hands was carried out on 80 subjects. Apart from a highly significant difference in performance between the sides in favour of the preferred hand, this task showed, as did the typewriting task, a marked positive correlation between the performance levels of the two sides. These findings are briefly discussed in relation to the training histories of the subjects.

INTRODUCTION

In a recent review of the literature on handedness and motor skill (Provins, 1967), it was concluded that there is little or no difference in the relative performance of the right and left sides on a variety of tasks which involve the execution of simple movements, but that significant differences do occur between the two sides where the serial organization of muscle activity is an important factor. These results were considered to be due primarily to the effects of differential training or usage of the limbs on the right and left sides of the body and to represent a special category of the well-recognized specificity of motor skills.

However, there are several implications of this differential training hypothesis which need to be tested, and the present report is a description of an experimental investigation of one of them. The purpose of the present series was to examine the performance of each side in two motor tasks which conformed to the description of a highly skilled activity (Provins, 1967), but which could be clearly distinguished on the basis of the amount of differential training previously given to the two sides. One task, namely typewriting, was chosen as being an activity which properly trained and experienced people normally only practised with both hands and that each hand could be considered therefore to have received approximately the same amount of usage. The other task, namely handwriting, was chosen as being an activity which people would have trained from infancy to undertake consistently with one particular hand and that this would therefore be thoroughly overtrained, whereas the other hand would have received no systematic practice on the task at all. The differential training hypothesis would predict no differences between the two sides by trained touch typists on the typewriting task but a clearly significant difference between the practised and unpractised sides in the handwriting task.

Three series of experiments will be described, two on the typewriting task and one concerned with handwriting performance.

TYPEWRITING EXPERIMENT: SERIES 1

Method

Two groups, each of 20 subjects were used. One group comprised trained touch typists with a mean age of 28 years (ranging from 17 to 50 years) and with a mean post-training experience of 4·6 years (and a minimum of 6 months). The other group with a mean age of 24·4 years (ranging from 17 to 32 years) had had no formal training at typing and were essentially non-typists. In the trained typists group there were two subjects who used their left hand for handwriting, whilst in the untrained group there were three left-handed writers. The subjects were obtained from university clerical, academic and technical staff.

---

* Now at Department of Psychology, Australian National University, Canberra.

The tasks consisted of seven short typing exercises. Details of these exercises are summarized in the appendix. The first exercise consisted of the repeated typing of 10 lines of a familiar sentence involving all of the letters of the alphabet. Exercises 2 and 3 (words) were constructed so that each involved one half of the keyboard only (according to Walmesley, 1958) and employed words made up from letters which occurred with approximately the same frequency in the test as they occur in normal English usage (Pratt, 1939). Exercises 4 and 5 (letters) comprised groups of five and six letters arranged in random order, each test again only involving the letters on one side of the keyboard. Although the groups of letters were different in the two exercises, they were arranged so that the order of fingering required was identical for the right and left hands. Finally, Exercises 6 and 7 (tapping) utilized the home keys only for each hand and were designed to require repetitive operation of the four fingers on the same side for each line of the test. Three lines of each of these two exercises employed a fingering sequence from index to little finger and three lines were typed in the reverse order, each fingering sequence being used on alternate lines.

For Exercise 1, subjects were asked to type in their usual way employing both hands as appropriate. For Exercises 2, 4 and 6, however, they were only allowed to use the right hand to type the letters, and for Exercises 3, 5 and 7, only the left hand, although no restriction was placed on the hand used to type the numbers or to operate the space bar and carriage return.

Subjects were told that although their results would be analysed for errors, the emphasis of the test was on speed of typing. Each exercise was timed separately and half of each group of 20 subjects carried out the tests in the order: Exercise 1, 2, 3, 4, 5, 6, and 7, while the other half completed the series in the order: Exercise 1, 3, 2, 5, 4, 7, 6.

Results

The time taken (in sec.) for each subject to complete each exercise and the number of errors made were recorded. The mean values and standard deviations for Exercises 2 to 7 are given in Tables I and II for the trained and untrained groups, respectively. (Exercise 1 was not used in the analysis of the results.)

TABLE I

COMPARISON OF MEAN TIMES (IN SEC.) AND ERRORS MADE BY TRAINED TYPISTS TO TYPE THREE DIFFERENT EXERCISES WITH EACH HAND (SERIES 1)

			Words		*Letters*		*Tapping*	
			Times	*Errors*	*Times*	*Errors*	*Times*	*Errors*
Mean values for right and left	R.	M.	111·6	8·8	123·2	12·4	127·8	13·2
		S.D.	31·11	5·95	29·12	8·30	32·58	8·97
	L.	M.	96·0	7·6	123·0	9·2	121·1	10·4
		S.D.	19·89	5·75	29·96	8·48	24·94	7·41
Comparison between R. and L.	All typists (n = 20)	*t*	4·23**	0·83	0·09	1·74	2·24*	1·74
		r	0·89**	0·42	0·94**	0·57*	0·93**	0·62*
	R.H. typists only (n = 18)	*t*	3·90**	0·62	0·19	1·75	2·14*	1·77
		r	0·90**	0·58*	0·93**	0·58*	0·93**	0·59*

* Signif. at 0·05 level.
** Signif. at 0·01 level.

TABLE II

COMPARISON OF MEAN TIMES (IN SEC.) AND ERRORS MADE BY NON-TYPISTS TO TYPE THREE DIFFERENT EXERCISES WITH EACH HAND (SERIES 1)

			Words		*Letters*		*Tapping*	
			Times	*Errors*	*Times*	*Errors*	*Times*	*Errors*
Mean values for right and left	R.	M.	279·2	2·2	292·5	2·7	161·3	8·5
		S.D.	94·62	2·29	67·55	1·93	39·39	9·39
	L.	M.	269·8	2·4	315·5	3·1	173·1	9·6
		S.D.	97·55	1·93	67·32	2·43	40·48	8·32
Comparison between R. and L.	All non-typists (n = 20)	*t*	1·37	0·35	3·81**	0·64	3·80**	0·71
		r	0·95**	0·29	0·92**	0·20	0·94**	0·70**
	R.H. non-typists only (n = 17)	*t*	0·81	0·25	3·28**	0·66	3·23**	0·95
		r	0·96**	0·33	0·91**	0·20	0·93**	0·66*

* Signif. at 0·05 level.
** Signif. at 0·01 level.

Using a paired comparison *t* test, only two significant differences were found between the performances of the right and left hands of trained typists (Table I). The left hand was significantly faster both at the speed of typing words ($p = 0{\cdot}01$) and the speed of repetitive tapping ($p = 0{\cdot}05$). This finding was still true when only the 18 right-handed subjects were considered by themselves. With the non-typists (Table II), the performances of the right hand were significantly better than the left hand on letter typing speed and the repetitive fingering sequence (tapping speed). These differences were still significant for the 17 right-handed non-typists. There were no significant differences between any of the error scores.

The results of product-moment correlation tests between the performances of the two sides are also given in Tables I and II. It can be seen that the times taken by typists on the right and left sides in each exercise are all significantly associated. The correlations between the errors made on each side are also positive in all cases and sometimes significantly so. Significant correlations between performance times on the two sides were also found with the non-typists, but the only significant correlations here for the error measures was on the tapping task.

A discussion of these results will be postponed until the results of the second experimental series has been described, as an important study reported by Birren and Botwinick (1951) has demonstrated that large age effects occur in speed of handwriting both words and digits, particularly after the age of about 40 years. Unfortunately, the present authors were unaware of this report until the first experimental study had been completed. As some of the trained typists used in this investigation were over 40 years of age, it was considered that the present typewriting results may in part be influenced by the age factor and to try to eliminate this possibly confounding influence, a repeat experiment was conducted on trained typists more closely matched for age with the control group.

TYPEWRITING EXPERIMENT: SERIES 2

Method

Twenty trained touch-typists who were members of the university staff were employed as subjects. They had a mean age of 20·2 years (ranging from 17–27 years) and an average of 4·1 years (with a minimum of 6 months) post-training experience. Two of the subjects used their left hand for handwriting.

The procedure, methods and typewriting exercises were identical with those reported in Series 1.

Results

The results of this repeat study were remarkably consistent with the findings of the previous experiment and are summarized in Table III.

TABLE III

Comparison of Mean Times (in sec.) and Errors made by Trained Typists to Type Three Different Exercises with Each Hand (Series 2)

			Words		*Letters*		*Tapping*	
			Times	*Errors*	*Times*	*Errors*	*Times*	*Errors*
Mean Values for right and left	R.	M.	96·93	5·90	106·91	9·71	115·80	12·00
		S.D.	35·9	4·3	29·2	6·66	22·7	21·46
	L.	M.	85·00	4·25	109·65	8·75	112·08	6·05
		S.D.	10·2	4·3	29·8	6·76	17·1	3·96
Comparison between R. and L.	All typists	*t*	4·79**	1·99	1·39	1·08	1·38	1·54
	(n = 20)	*r*	0·84**	0·63*	0·96**	0·81**	0·85**	0·83**
	R.H. typists	*t*	4·25**	1·57	1·15	1·11	1·34	1·68
	only (n = 18)	*r*	0·84**	0·64*	0·96**	0·92**	0·51	0·86**

* Signif. at 0·05 level.
** Signif. at 0·01 level.

Once again the left hand of these trained typists tended to be superior in performance, but the only difference which reached significance in this experiment was in the speed of typing words. This was still true when only the 18 right-handed subjects were considered by themselves. As with the first experiment, performance on the right and left sides is significantly associated. The correlations between the two sides for all time and error measures were significantly positive, except one.

The results of this second study, where age is controlled, and the initial study are so alike that any confounding influence due to age may be considered to be relatively insignificant, and the results of the two series will therefore be discussed together.

Discussion

It is worth noting that no significant differences in errors were recorded between the performances of the two sides on any exercise by either the trained typists or the group of non-typists. It is therefore fair to say that any differences recorded between the two sides in speed of performance were not at the expense of accuracy.

In evaluating the relative speed of performance on the two sides, quite clear-cut differences were found between the typists and non-typists on each of the three exercises. For the letter and tapping exercises the non-typists showed a highly significant difference between hands in favour of the right side whereas the trained subjects showed either no differences on these exercises or a significant difference in favour of the left hand. On the word exercise, however, no differences between sides were displayed by the untrained subjects whereas the touch typists showed highly significant differences in favour of the left hand.

It is suggested that the most likely explanation for the different patterns of performance shown by the typists and non-typists may be sought in their different training histories. For the letter and tapping exercises, there were little or no differences found between the sides for the touch-typists as was predicted, due presumably to the equal amount of prior practice with each hand. The slight tendency (but nevertheless significant one in Series 1) for trained typists who are otherwise right-handed to perform better with the left hand in the tapping exercise is perhaps a little surprising, but may be due to the fact that the left hand receives somewhat more practice than the right in normal typewriting work. The allocation of letters between the two hands means that in typing English with the standard keyboard, the left hand is required to type over 56 per cent. of the total number of strokes performed in any given text (Hoke, 1922; Dvorak, 1943). Furthermore, it appears true to say that in normal typewriting, the left hand also receives somewhat greater practice in the letter sequences which are used to make up complete words. Certainly in devising Exercises 2 and 3, it was found to be much easier to construct complete words for typing with the left hand than for the right. Thus the word exercises probably presented the experienced typist with more familiar fingering sequences and the tapping exercises more practised movements on the left side than the right. In contrast, the serially random nature of the letter exercises probably minimized the extent to which previous training of this sort could be employed.

No such considerations concerning previous training or usage on the task apply to the non-typists. For these subjects, it seems likely that the lack of specific training on typewriting forced them to try to transfer to this task whatever other similar skills they possessed, and as these were better developed for the right hand for most people, they performed these exercises better with that hand. For the word exercise, a similar explanation may apply but with an additional consideration, for whereas the letter and tapping exercises were designed to be strictly comparable on the two sides for fingering sequences and frequency of usage for any one finger, the same balancing effect could not be achieved with the word test. It is possible, therefore, that this exercise was somewhat easier for the left hand for all subjects and not only enabled the typists to do better with this side, but reduced the effectiveness of transfer of right-handed training to this task in the case of the non-typists. It is perhaps worth pointing out however that in accounting for the performances of the different groups of subjects on the various exercises, the dependence on vision would have been considerably less for the touch-typists than for the non-typists. It is therefore possible that eye-dominance may have contributed in some way to the effects reported here although only further research can determine the importance of this particular variable.

One further important finding in the present experimental series concerns the high degree of association between the level of performance achieved with the right side and the left by both the trained typists and untrained subjects as shown by the significant correlation values in each Table. This indicates that irrespective of whether extensive training is received on both sides or none at all, subjects who do well with one hand will tend to do well with the other on this particular task. Whether this is simply a reflection of the degree of motivation of each subject or of differences in some basic motor capacity is somewhat conjectural, and will be considered again in the General Discussion later, but the effect seems to be particularly strong and consistent. It was therefore thought worthwhile to make a similar comparison between sides in a motor task thoroughly over-learned with one hand and quite unpractised with the other and this forms the substance of the next part of this study.

HANDWRITING EXPERIMENT: SERIES 3

Method

Eighty subjects were tested on a simple handwriting task. The sample consisted of the two groups of 20 trained typists used in the experiments in Series 1 and 2. The remaining 40 subjects, of mean age 22·0 years (ranging from 17 to 32 years) were non-typists including 19 from the control group used in Series 1 and 2.

The task involved writing out the alphabet in pencil from a to z in small letters (as if it were one word) six times as quickly as possible. The subject was allowed to write the alphabet twice for practice purposes with each hand and he was told that while legibility was necessary, the emphasis was on speed of execution. The subject was timed by means of a stop watch and after the exercise had been completed with the preferred hand, he was told how long he had taken and then asked to repeat what he had just done with the other (non-preferred) hand.

Results

The *t* values and correlations for the comparisons between the means of the preferred and non-preferred hands for each group of subjects are summarized in Table IV. From this table it can be seen that for both the typists and non-typists, a marked superiority of the preferred hand in speed of handwriting was recorded.

TABLE IV

COMPARISON OF MEAN TIMES TAKEN (IN SEC.) BY TYPISTS AND NON-TYPISTS IN USING PREFERRED AND NON-PREFERRED HANDS TO PERFORM HANDWRITING TASK (SERIES 3)

		Typists		
		Series 1 ($n = 20$)	*Series 2* ($n = 20$)	*Non-typists* ($n = 40$)
Preferred hand	M.	77·9	84·78	85·19
	S.D.	9·48	12·35	12·16
Non-preferred hand	M.	157·9	171·14	171·63
	S.D.	33·74	29·14	37·74
	t	11·73**	13·72**	17·13**
	r	0·54**	0·29	0·60**

** Signif. at 0·01 level.

For this task, correlations of 0·54 and 0·29 were found between the performances of the two sides for the typists of Series 1 and 2 respectively and a corresponding correlation of 0·60 recorded for the control group of non-typists.

Discussion

The highly significant differences between the two sides for all subjects is, of course, as predicted, and may be ascribed to the previous differential training of the two hands on this task.

The positive correlation between the level of performance achieved with each hand parallels the results found for the typewriting task. However, the effects here are not so marked, although two out of the three groups showed highly significant correlations between the two sides. These effects will be considered together with the typewriting results in the General Discussion.

General Discussion

It has been shown previously that in simple motor tasks, such as tests of muscle strength (Roberts, Provins and Morton, 1959) and accuracy of reproduction of pressure (Provins, 1956), no differences in performance between the two sides may be recorded. Similarly, it has been shown that in motor tasks demanding considerable serial organization of muscle activity, no differences between the two sides are evident if no training or practice has been possible on either side (Provins, 1956). On this latter test of skill (tapping with the big toe), training or practice was presumed to be equal on the two sides and minimal. In the present typewriting task for the touch typists, previous training was considered to be the same for each hand and maximal (i.e. overlearned), with results again showing little or no difference between performances on the two sides. Where differences occurred, these were in favour of the (mainly) non-preferred hand for handwriting and were most likely to be due to differences in difficulty of the exercises for the right and left hands and/or the slightly greater practice which the left hand receives in normal typing.

In tasks requiring motor skill and where differences between the two sides in previous training is known or can be assumed to have occurred, significant differences in performance between right and left limbs have been demonstrated in earlier researches (Provins, 1956; 1958) and again in the handwriting experiment reported here.

A particularly interesting finding in the present series of studies, however, concerns the generally high degree of association between the level of performance achieved on the two sides in both the typewriting and handwriting tasks. It is, of course, possible that this simply reflects differences between subjects in their degree of motivation, but there are other reasons for believing that this may not be the only or even the most likely explanation. This finding could well be explained by postulating that different individuals have a different capacity to acquire motor skill so that if training is held constant for a given group of subjects (i.e. they all "overlearned" the task with both hands or one hand only, as in the typewriting and handwriting experiments respectively), those subjects who did well with one hand would be expected to do well with the other. It has been shown previously that dull and mentally deficient children not only obtain low scores on tests of motor proficiency (Sloan, 1951), but also tend to possess poorly-developed handedness. There is also clinical evidence to suggest that in children of normal intelligence who are suffering severe motor deficiencies in the absence of any other neurological disturbance, handedness may not be clearly differentiated (see for example, Walton, Ellis and Court, 1962). In other words, there appears to be a reasonable basis for suggesting that a minimum level of general intelligence and motor capacity is necessary for the acquisition of motor skill and that without these, handedness is also unable to develop. With an adequate capacity for the acquisition of motor skill, the level of performance achieved by an individual in a motor task appears to depend on the extent of his training on this or similar tasks and that in so far as the training of the two sides has been dissimilar, then the levels of performance achieved on the two sides are dissimilar.

References

Birren, J. E. and Botwinick, J. (1951). The relation of writing speed to age and to the senile psychoses. *J. Consult. Psychol.*, **15**, 243–9.

Dvorak, A. (1943). There is a better typewriter keyboard. *Nat. bus. educ. Quart.*, **12**, 51–58, 66.

Hoke, R. E. (1922). Improvement of speed and accuracy in typewriting. *Johns Hopkins Studies in Education*, No. 7, 1–42.

PRATT, F. (1939). *Secret and Urgent. The Story of Codes and Ciphers.* London: Robert Hale.
PROVINS, K. A. (1956). Handedness and skill. *Quart. J. exp. Psychol.*, **8**, 79–95.
PROVINS, K. A. (1958). The effect of training and handedness on the performance of two simple motor tasks. *Quart. J. exp. Psychol.*, **10**, 29–39.
PROVINS, K. A. (1967). Motor skills, handedness and behaviour. *Aust. J. Psychol.*, **19**, 137–150.
ROBERTS, D. G., PROVINS, K. A., and MORTON, R. J. (1959). Arm strength and body dimensions. *Human Biol.*, **31**, 334–43.
SLOAN, W. (1951). Motor proficiency and intelligence. *Amer. J. Ment. Defic.*, **55**, 374–406.
WALMESLEY, W., and E. (1958). *Pitman's Commercial Typewriting.* London: Pitman 4th edition.
WALTON, J. N., ELLIS, E., and COURT, S. D. M. (1962). Clumsy children: A study of developmental apraxia and agnosia. *Brain*, **85**, 603–12.

Manuscript received 19*th February*, 1968.

APPENDIX

Exercise 1

1. The quick brown fox jumps over the lazy dog.
2. The quick brown fox jumps over the lazy dog.
(Sentence repeated 10 × = 10 lines altogether)

Exercise 2

1. inulin imply hobo phylum hooky pink humbly bloom hunk pinion hilum pony
2. inulin imply hobo phylum hooky pink humbly bloom hunk pinion hilum pony
(12 words repeated 5 × = 5 lines altogether)

Exercise 3

1. vaster screw stag access draft fact scared dwarf vest dagger frets weft
2. vaster screw stag access draft fact scared dwarf vest dagger frets weft
(12 words repeated 5 × = 5 lines altogether)

Exercise 4

onmjlu yikhp pknoyl imjhu lujkmp hnioy jmhkop uilny hlopjm uiykn
iymjhu pknlo polyiu njmhk mnpjil khyuo mpkjou ihlyn unohyp iljkm
(Both lines repeated in reverse order = 4 lines altogether)

Exercise 5

wcxfsr tedgq qdcwts exfgr srfdxq gcewt fxgdwq resct gswqfx retdc
etxfgr qdcsw qwster cfxgd xcqfes dgtrw xqdfwr egstc rcwgtq esfdx
(Both lines repeated in reverse order = 4 lines altogether)

Exercise 6

;lkj ;lkj ;lkj ;lkj ;lkj ;lkj ;lkj ;lkj ;lkj ;lkj ;lkj ;lkj ;lkj ;lkj ;lkj ;lkj ;lkj ;lkj
jkl;jkl;jkl;jkl;jkl;jkl;jkl;jkl;jkl;jkl;jkl;jkl;jkl;jkl;jkl;jkl;jkl;jkl;
(Both lines repeated alternately 3 × = 6 lines altogether)

Exercise 7

asdfasdfasdfasdfasdfasdfasdfasdfasdfasdfasdfasdfasdfasdfasdfasdfasdfasdfasdfasdf
fdsafdsafdsafdsafdsafdsafdsafdsafdsafdsafdsafdsafdsafdsafdsafdsafdsafdsafdsafdsa
(Both lines repeated alternately 3 × = 6 lines altogether)

SHORTER ARTICLES AND NOTES

INVOLVEMENT OF NECK PROPRIOCEPTIVE SYSTEM IN VISUAL AFTER-EFFECT FROM PROLONGED HEAD TILT

BY

R. H. DAY and N. J. WADE

From the Department of Psychology, Monash University, Melbourne

Judgements of object orientation after prolonged head tilt differ relative to those made before tilt. This visual spatial after-effect was investigated with lateral head tilt relative to the upright and supine body. The data from two experiments indicate that the magnitude and direction of the visual after-effect does not differ under the two body postures. It is concluded that the proprioceptive system of the neck, as opposed to the otolith system of the utricle, is involved in the after-effect.

INTRODUCTION

In the upright posture a dimly lighted line in an otherwise dark or featureless environment can be accurately adjusted to the vertical (Mann, Berthelot-Berry and Dauterive, 1949; Neal, 1926). With head upright following a 2 or 3 min. period of lateral head tilt with eyes closed apparent orientation judgements shift in the direction of prior tilt relative to pre-tilt judgements (Day and Wade, 1966). The development and dissipation functions for this visual after-effect are exponential (Wade and Day, 1968). The after-effect as a function of angle of tilt is opposite in direction to the angular function for the effect which occurs during tilt. It is assumed that the systems mediating the effect, when stimulated for prolonged periods, undergo adaptive change and give rise to the oppositely directed after-effect.

Head tilt stimulates otolith receptors in the labyrinthine utricle and the proprioceptive (joint and muscle) systems of the neck. The otolith organs respond to that component of gravity acting laterally, at right angles to the longitudinal axis of the head (Trincker, 1962). The magnitude of this component increases as a function of tilt with consequent increase in the displacement of the otoliths relative to the macular surface of the utricle. If, however, the head is tilted while the observer is supine the force acting on the otolith organs remains constant and no displacement occurs. Thus, occurrence of a visual after-effect following head tilt in a supine posture would indicate involvement of a sensory system other than that associated with the utricle, namely the neck proprioceptive system.

EXPERIMENT I

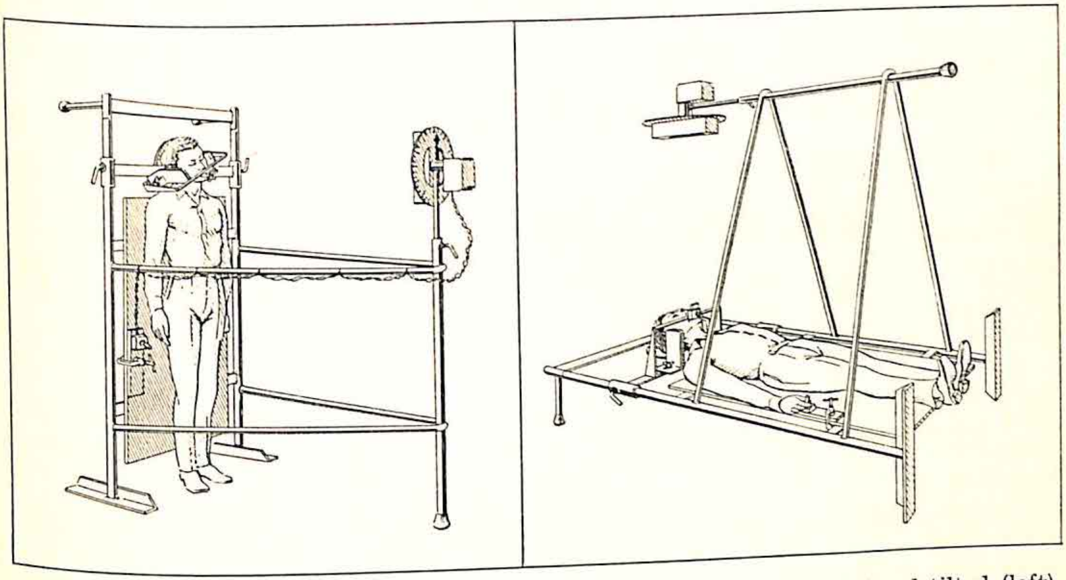
In previous studies of the visual after-effect from prolonged head tilt (Day and Wade, 1966; Wade and Day, 1968) the subject's task was to adjust a bar of light in an otherwise dark room to the vertical while seated upright. Orientation judgements with a supine subject cannot be made in terms of verticality since this is not a relevant orientation. Consequently, the instructions in the present experiments were for adjustment of the bar to alignment with the body axis. Preliminary observations had shown that there is no difference between alignments to the vertical and to the body axis with the subject upright.

Method

Subjects. Twelve members of Monash University participated as subjects. Nine were male and three female.

Apparatus. Lateral head tilt was controlled by a projecting rectangular frame pivoted on a crossbar between rigid steel tubes to which adjustable head supports and individual dental composition bite-boards were attached. The angle of head tilt could be measured to the nearest 0·25° by means of a protractor scale. The visual stimulus object was a 150 mm. × 1 mm. line of light (0·15 mlam.) which could rotate in a plane fronto-parallel to the subject. The stimulus object was supported on a larger projecting frame (also attached to the steel tubes) 100 cm. from the subject, who could adjust its orientation by operating a switch which activated a 1 r.p.m. synchronous reversible motor. The experimenter could adjust the line manually, and read its orientation to the nearest 0·1°. The whole apparatus could be rotated through 90° to accommodate either an upright or a supine subject (Fig. 1), and was situated in a dark room.

FIGURE I



Experimental arrangement showing subject in upright posture with head tilted (left) and supine with head and body aligned (right).

Procedure. The subject made five adjustments of the bar (from starting positions 0°, 5° and 10° clockwise and anti-clockwise of the body axis) to apparent body axis prior to 2 min. lateral 30° head tilt. During the head tilt the subject's eyes remained closed. A further adjustment was made upon returning the head to alignment with the body from one of the above starting positions. A single post-tilt adjustment was used since the after-effect dissipates rapidly on returning the head to the normal position (Wade and Day, 1968). The difference between the mean of the five pre-tilt adjustments and the single post-tilt adjustment was the measure of the after-effect. Each subject was tested with right and left head tilt in both upright and supine positions in balanced order. An

TABLE I

MEAN AFTER-EFFECTS (DEGREES) AND VARIANCES FOR UPRIGHT AND SUPINE POSTURES

	Experiment I	
	Upright	*Supine*
Mean	1·773	1·936
Variance	0·718	7·443
	Experiment II	
	Upright	*Supine*
Mean	1·883	2·145
Variance	0·296	0·842

interval of 5 min. between trials ensured dissipation of the after-effect before commencement of the subsequent trial (Wade and Day, 1968).

Results. Since there were no significant differences between the magnitudes of the after-effects from right and left head tilt ($t = 0{\cdot}767$, $p > 0{\cdot}05$ for upright, $t = 0{\cdot}207$, $p > 0{\cdot}05$ for supine) the two were combined for each posture. The mean after-effects for the two postural conditions are shown in Table I together with the variances. Each mean was significantly different from zero ($t = 7{\cdot}131$, $p < 0{\cdot}001$ for upright, $t = 2{\cdot}350$, $p < 0{\cdot}05$ for supine).

There was no significant difference between the after-effects for the upright and supine postures ($t = 0{\cdot}152$, $p > 0{\cdot}05$).

EXPERIMENT II

The variance for the supine condition is large relative to that for the upright with twelve untrained subjects. In an attempt to reduce this variance difference a second experiment was conducted using a small number of subjects highly trained in the task of adjusting the light bar to the body axis.

Method

Subjects. Four male graduate members of Monash University Psychology Department were subjects. Each subject had participated in a number of previous experiments involving visual judgements of orientation.

Apparatus. The apparatus was the same as that used in Experiment I.

Procedure. The procedure was essentially the same as that used in Experiment I except that 20 after-effect measures were obtained for each subject, 10 while upright and 10 while supine. Five sessions were required for each subject.

Results. The after-effects from right and left head tilt were combined and the mean after-effects for the two postures together with the relatively small variances are also shown in Table I. Both means were found to be significantly different from zero ($t = 6{\cdot}259$, $p < 0{\cdot}01$ for upright, $t = 4{\cdot}072$, $p < 0{\cdot}05$ for supine) but not significantly different from each other ($t = 0{\cdot}482$, $p > 0{\cdot}05$). Thus data from the second experiment confirm those of the first.

DISCUSSION

Occurrence of a visual after-effect following prolonged head tilt while supine cannot be attributed to otolith function. Further, the observation that there is no difference in the magnitude of the after-effects between the upright and supine positions strongly suggests that the otolith organs do not contribute to the phenomenon in the former posture. A growing body of behavioural evidence indicates non-adaptation of the otolith organs (Clark and Graybiel, 1962; 1966; Graybiel, 1952). These data provide additional support for this conclusion in that prolonged stimulation of the neck and otolith systems during head tilt in the upright posture produces an after-effect of the same magnitude as from stimulation of the neck system alone in the supine position.

These results accord with neurophysiological (Cohen, 1961; McCouch, Deering and Ling, 1951) and clinical (Cope and Ryan, 1959; Gray, 1956; Weeks and Travell, 1955) data which show that the kinaesthetic systems of the cervical joints (and possibly the deep neck musculature) play an important role in visual spatial orientation. It is reasonable to suppose, therefore, that this system undergoes adaptation during prolonged head tilt and in consequence visual orientation is affected.

REFERENCES

CLARK, B., and GRAYBIEL, A. (1962). Visual perception of the horizontal during prolonged exposure to radial acceleration on a centrifuge. *J. exp. Psychol.*, **63**, 294–301.

CLARK, B., and GRAYBIEL, A. (1966). Perception of the visual horizontal in normal and labyrinthine defective observers during prolonged rotation. *Amer. J. Psychol.*, **79**, 608–12.

COHEN, L. A. (1961). Role of eye and neck proprioceptive mechanisms in body orientation and motor coordination. *J. Neurophysiol.*, **24**, 1–11.

COPE, S., and RYAN, G. M. S. (1959). Cervical and otolith vertigo. *J. Laryng.*, **73**, 113–20.

Day, R. H., and Wade, N. J. (1966). Visual spatial aftereffect from prolonged head tilt. *Science*, **154**, 1201–2.

Gray, L. P. (1956). Extralabyrinthine vertigo due to cervical muscle lesions. *J. Laryng.*, **70**, 352–60.

Graybiel, A. (1952). Oculogravic illusion. *Arch. Ophthal., Chicago*, **48**, 605–15.

Mann, C. W., Berthelot-Berry, N. H., and Dauterive, H. J. (1949). The perception of the vertical: I. Visual and non-labyrinthine cues. *J. exp. Psychol.*, **39**, 538–47.

McCouch, G. P., Deering, I. D., and Ling, T. H. (1951). Location of receptors for tonic neck reflexes. *J. Neurophysiol.*, **14**, 191–5.

Neal, E. (1926). Visual localization of the vertical. *Amer. J. Psychol.*, **37**, 287–91.

Trincker, D. (1962). The transmission of mechanical stimulus into nervous excitation by the labyrinthine receptors. *Symp. Soc. Exp. Biol.*, **16**, 289–314.

Wade, N. J., and Day, R. H. (1968). Development and dissipation of a visual spatial aftereffect from prolonged head tilt. *J. exp. Psychol.*, **76**, 439–43.

Weeks, V. D., and Travell, J. (1955). Postural vertigo due to trigger areas in the sternocleidomastoid muscle. *J. Pediat.* **47**, 315–27.

Manuscript received 28th October, 1967.

AN INQUIRY INTO INAPPROPRIATE CONSTANCY SCALING

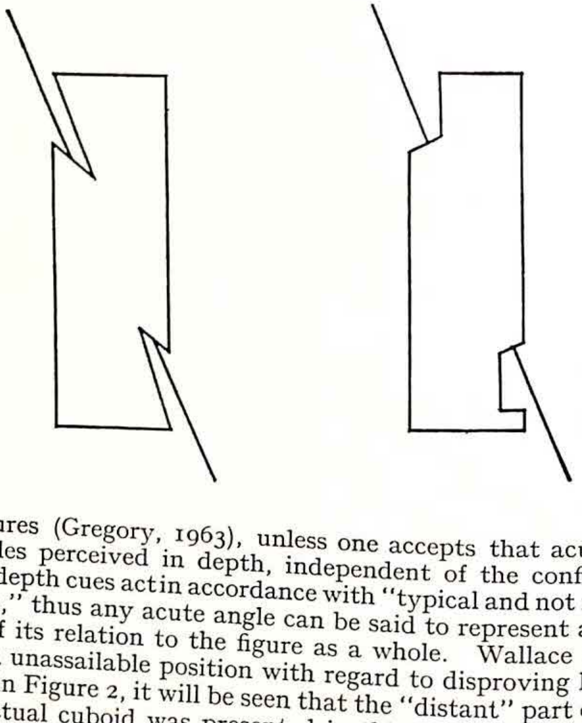
BY

H. J. JEFFREY

From Department of Psychology, University of Sheffield

The importance of acute or obtuse angles in the Poggendorff illusion may be seen in Figure 1. Where the angles are reversed the illusion is reversed; where the angles are removed the illusion disappears. It is difficult to interpret these effects in terms of primary depth features (Gregory, 1963), unless one accepts that acute angles are interpreted as right angles perceived in depth, independent of the configuration. Gregory argues that primary depth cues act in accordance with "typical and not necessarily apparent depth configurations," thus any acute angle can be said to represent a right angle seen in depth irrespective of its relation to the figure as a whole. Wallace (1966) suggests this leaves Gregory in an unassailable position with regard to disproving his hypothesis.

FIGURE 1



With the cuboid in Figure 2, it will be seen that the "distant" part of the figure appears expanded. If an actual cuboid was presented in this way there would be no apparent contraction of the distant part, it being compensated by perceptual enlargement. Given only certain cues to depth, but in the absence of apparent depth, perceptual enlargement remains. This is surely inappropriate constancy scaling and the cues that function independent of apparent depth are surely "primary depth features." Furthermore, the distortion approximates to what would be predicted from the expansion of acute angles (the outside faces being an example of the Zollner figure).

If the same figure is viewed as an open book (Fig. 3), the opposite end appears expanded, supporting Gregory in as much as the acute angles appear not necessarily to follow the single rule of expansion, but rather act in accordance with their function as depth cues. However, the Zollner illusion although reduced appears to conflict with the general configuration, and is certainly not reversed in the same way. It might be maintained that the "typical" function of primary depth cues counteract even their appropriateness, but these additional qualifications must reduce the value of the proposition. Certainly two definitions of primary scaling appear incorrect (Gregory, 1966). It is subject to change by set, and it does change with the interpretation of depth in ambiguous figures presented on textural planes.

With illusions of extension, e.g. Müller-Lyer and Dumbell, it is difficult to accept size constancy as an explanation once it is realized that nearly any two-dimensional extensions

FIGURE 2

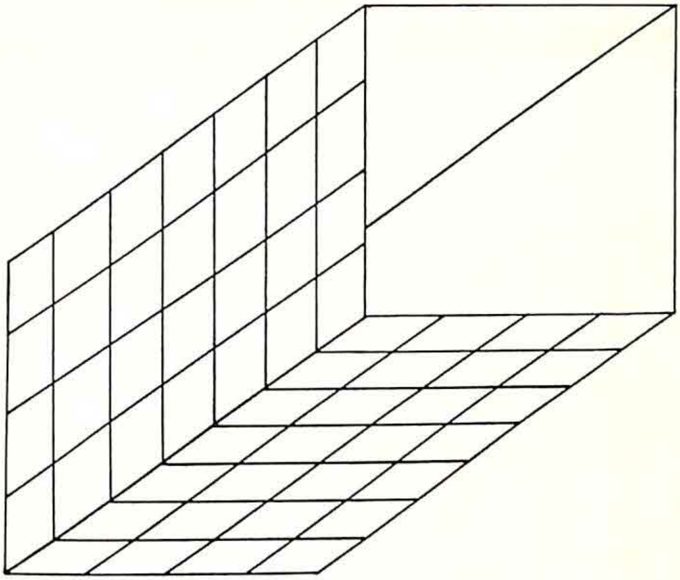


FIGURE 3

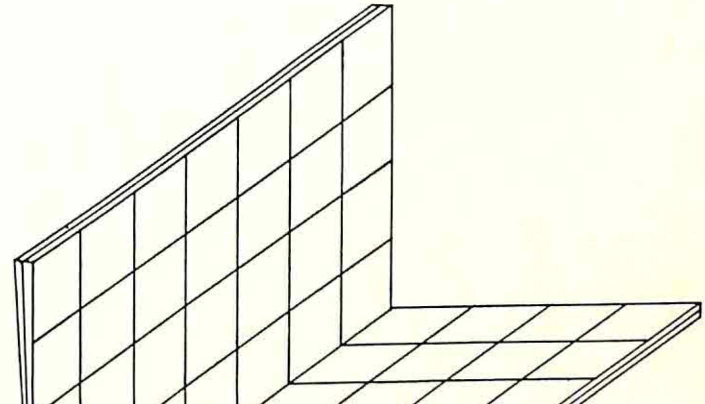
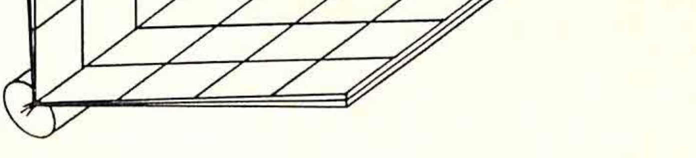


FIGURE 4



to each end of a line or linear area increase its perceptual length. Thus one may create a figure in which depth cues such as masking or perspective are absent without reducing the illusory effect (see Fig. 4, comparing a with b). What does seem significant is the area of the extensions. Thus if they are expanded horizontally (e.g. from squares to elongated rectangles) the area between them reduces as a function of the expansion until the illusion of overestimation disappears. If they are extended vertically, the horizontal distance between them remains overestimated, i.e. the illusion persists.

The above findings are substantiated by a series of experiments incorporated in a dissertation submitted to the Department of Psychology, Sheffield University.

REFERENCES

GREGORY, R. L. (1963). Distortion of visual space as inappropriate constancy scaling. *Nature*, **199**, 678–9

GREGORY, R. L. (1966). *Eye and Brain*. London: Weidenfeld & Nicolson. Pp. 31–63.

WALLACE, G. K. (1966). Optical illusions. *Nature*, **209**, 327–8.

Manuscript received 8th January, 1968.

THE TIME-COURSE OF PREPARATION WITH REGULAR AND IRREGULAR FOREPERIODS

BY

PAUL BERTELSON and FRANÇOISE TISSEYRE

From the Laboratoire de Psychologie, Université Libre de Bruxelles

A previous experiment (Bertelson, 1967) had shown that the temporal information brought by a warning signal affected RT even after very short foreperiods (FPs). The present experiment was carried out to examine whether this result was contingent on the predictability of the FP. After a 5 sec. waiting delay, the subject heard a warning click which was followed after a predictable (regular procedure) or unpredictable FP (irregular procedure) by the visual signal calling for a choice reaction. The range of FPs was 0–300 msec. again. The time course of the adjustments triggered by the click was found to be similar under both procedures. The main conclusion is that a shift from preparation to reaction can occur at any time and need not be programmed before preparation is started.

INTRODUCTION

In a previous experiment (Bertelson, 1967), the time course of preparation was studied by having the stimulus calling for a choice reaction preceded at different short foreperiods (FPs) by an auditory warning signal. The regular procedure was used for the presentation of FPs: each duration was always presented for a run of trials, allowing the subject to adapt as well as he could to that particular value. Effective preparation was found even after very short FPs like 50 msec. This was taken as evidence that a signal can be attended to, and used as a time cue, without starting a refractory period of the order of duration usually found after signals calling for an immediate reaction. It led to the suggestion that preparation is not a process which engages central decision mechanisms in an intermittent way, or, in other words, that when preparation is under way, the analysis of the stimulus need not be delayed till it is completed.

In a preliminary comment on this result (Bertelson, 1966), the question was asked whether it was contingent on the predictability of the FP: to be able to switch without delay from preparation to reaction, is it necessary to know beforehand, i.e. at the moment preparation is started, when the stimulus will come?

In the following experiment, the effects of the same FPs will be compared under a regular and under an irregular presentation procedure. The time of occurrence of the warning signal itself was made highly unpredictable, as in the previous experiment, but this time only by the use of a long constant waiting delay. The experimental situation differed on details only, such as the fingers used to respond, or the installation of the subject in an isolated cubicle. The main differences between the two experiments are the subject populations (students instead of sailors) and the fact that they now worked under monetary incentive, with an explicit payoff matrix.

METHOD

Apparatus. The reaction stimuli were provided by two neon pea-lamps mounted at an horizontal distance of 2 cm. (between centres) behind a vertical panel of dark-blue glass, at about eye-level. A third lamp was used to signal the beginning of each trial. It was situated 10 cm. below the stimulus lamps. Responses were given by pressing one of two horizontal keys, placed 10 cm. apart, and on which the two index fingers rested permanently (critical pressure: about 100 gr.). The click providing the warning signal was delivered through earphones. Its intensity, measured with a Marconi (TF 2160) Attenuator was about 75 dB. above threshold. The subject was installed in a cubicle isolated from the noise of the programming equipment, so that no masking had to be applied. The experimenter talked to him over an interphone.

A trial started when the experimenter activated a starting switch. An *experimental trial* consisted of the following sequence: (*a*) 100 msec. flash in the lower lamp; (*b*) 5 sec. waiting delay; (*c*) warning click; (*d*) foreperiod; (*e*) onset of one of the two stimulus lamps, which then stayed on till the start of the following trial. On a *control trial*, no click was applied, and the stimulus came immediately at the end of the waiting delay.

The waiting delays and the FPs were controlled by timing units using unijunction transistors, which gave an accuracy better than 1 per cent. RTs were measured on a dekatron timer with a 1,000 cps. time-base. The sequence of stimuli and also, in the irregular condition, of FPs were programmed by a Creed 8-bank punched tape reader.

Subjects. Eight male university students participated in five sessions each, one in the morning and one in the afternoon of successive days. They were paid for their services at a fixed hourly rate, plus a bonus depending on the number of "fast" reactions (see below) and of errors, one error being worth five fast reactions.

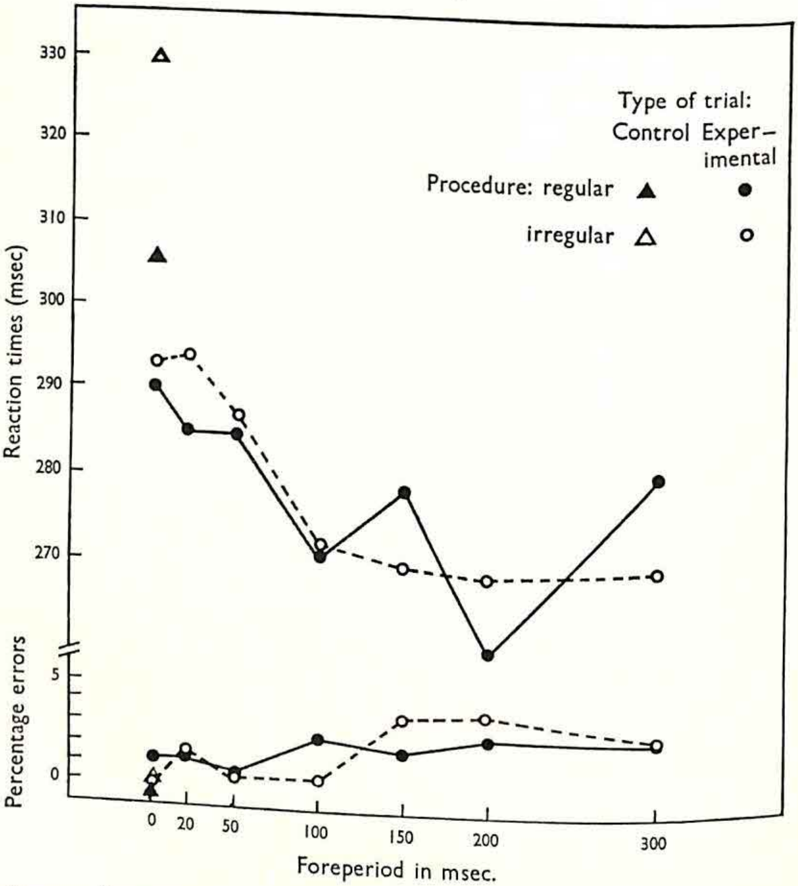
Procedure. The *regular condition* involved the same procedure as the previous experiment. At each session, one block of trials was run with each of the seven FP durations: 0, 25, 50, 100, 150, 200 and 300 msec., together with one block of control trials. A block consisted of 12 errorless trials, of which the last 10 only provided the data for the analysis. The order of the eight blocks was organized in latin squares.

Under the *irregular* condition, experimental trials with the different FP durations and control trials were administered in random order, with equal relative frequencies. 150 trials were given at each session.

Session 1 was devoted to practice under the two conditions. The data were gathered during sessions 2–5. For each subject two of these sessions were run under the regular and two under the irregular procedure, in alternated order.

Knowledge of results was given after each trial, using the simple code: "fast" or "slow," meaning faster or slower than the mean RT with the same FP on the previous session. In case of error, no information regarding speed was communicated.

FIGURE 1



Reaction time and percentage errors as functions of foreperiod duration, with procedure (regular vs. irregular) as parameter.
RTs: averages over sessions and subjects of median correct RTs per subject, session and FP.
Errors: averages over sessions and subjects of percentages per subject, session and FP.

RESULTS

The analysis was based on median RTs per subject, session and FP. Averages over subjects and sessions are given in the figure.

For the experimental trials (trials with warning click), no systematic difference between the two procedures, neither in the general level nor in the evolution with increasing FP, is observed. This is confirmed by the results of a variance analysis applied to the experimental trials medians. The interaction between FPs and procedures is non-significant. So is the FP × subjects interaction, but the FPs produce a highly significant variation ($F = 12{\cdot}2$ with 6 and 42 *d.f.*, $p = 0{\cdot}001$). The fact, apparent in the figure, that the curve is much smoother for the irregular procedure can probably be attributed to the better control of within-sessions order effects which this procedure entails.

The control trials give longer RTs than the corresponding experimental trials at FP = 0 in both procedures. The difference is in both cases shown by every individual subject, and is thus significant beyond $p = 0{\cdot}01$ by a simple sign test. The mean difference is much bigger (38 msec.) under the irregular than under the regular (16 msec.) procedure. In fact, there is a large (24 msec.) difference between the control RTs under the two procedures. The difference is again shown by every subject.

The percentages errors are also shown in the Figure. Practically no errors were observed on control trials, and very few on experimental trials.

DISCUSSION

(1) The main conclusion is that the time course of the adjustments triggered by a warning signal is the same whether the FP is predictable or not. Foreknowledge of the moment of occurrence of the stimulus is not necessary for efficient preparation to take place. If the analysis proposed in the introduction is accepted, this means that the switching from preparation to reaction need not be pre-programmed, but can occur at any moment, the RT depending of course on the amount of time spent preparing.

(2) The results must not be taken as contradicting the well-established principle that RT depends on time-uncertainty. The independence of preparation course from foreknowledge must of course be expected only over a very short range of FPs, like the one considered in the present experiment. With such short FPs, it is presumably advantageous always to start preparation immediately. If longer FPs, for which delayed preparation would be more beneficial, were included, the subject would have to choose a strategy for starting preparation, and range effects of the sort demonstrated by Drazin (1961) might be expected to occur in the irregular condition.

(3) If foreknowledge of the duration of the FP seems to be irrelevant, information regarding the fact that there will be a warning click or not is very important. This is implied by the difference between the RTs on the no-click control trials under both conditions. The reaction is definitely faster when it is known beforehand that no click will occur. When a click is expected and does not occur, it takes some extra time to start reacting on the basis of the visual stimulus only.

(4) It has been suggested by one referee that some tendency to expect the click might have occurred also on the control trials of the sessions run under the regular procedure, where it would have been transferred from the preceding blocks of experimental trials. This would account for the longer RTs observed on these control trials, compared with the experimental trials with FP = 0. A transfer effect must normally be felt less and less on successive trials of a control block. This reasonable prediction, which was made by the same referee, has been tested by computing mean RTs, corresponding to successive control trials, over the eight subjects No trend for RTs to get shorter was observed. This leaves only the two (not mutually exclusive) explanations suggested in the previous paper: (*a*) a central warning effect of the peripherally simultaneous click, permitted by differences in conduction times, and (*b*) an immediate facilitative effect of the click on the reaction to the light.

(5) The reduction of RT with increasing FP is of the same order of magnitude as in the previous experiment, but it is now obtained without the important increase in the proportion of errors that accompanied it. In the irregular condition, the whole reduction in RT is practically obtained after 100 msec. and the corresponding percentage errors is negligible (0·5 per cent.). The possibility that the increase in speed was due solely to a change in strategy can now safely be excluded, and in a much more satisfactory way than through the "correction for guessing" which had been applied. The smaller number of

errors in the present experiment is probably an effect of the use of a payoff matrix which discouraged wrong reactions.

(6) The slope of the RT–FP curve has now been found in two separate experiments to be about 0·25, over the 0–100 msec. interval. This can be interpreted as a sort of trade-off function between preparation and reaction. The efficiency of preparation is only one-quarter that of reaction. It means that if one adopts the simple view (e.g. Bertelson, 1961; Le Ny, 1956) that preparation consists of "doing in advance" some of the operations involved in reacting, one must add that they are done much more slowly.

This work has been partly supported by the Belgian Fonds National de la Recherche scientifique under the form of a "Crédit aux Chercheurs." The data were analysed under contract 612 with the "Fonds de la Recherche fondamentale collective."

REFERENCES

BERTELSON, P. (1961). Sequential redundancy and speed in a serial two-choice responding task. *Quart. J. exp. Psychol.*, **13**, 90–102.

BERTELSON, P. (1966). Central intermittency twenty years later. *Quart. J. exp. Psychol.*, **18**, 153–63.

BERTELSON, P. (1967). The time course of preparation. *Quart. J. exp. Psychol.*, **19**, 272–9.

DRAZIN, D. H. (1961). Effects of foreperiod, foreperiod variability and probability of stimulus occurrence on simple reaction time. *J. exp. Psychol.*, **62**, 43–50.

LE NY, J. F. (1956). Le temps de réaction motrice simple considéré comme un indicateur psychologique. *La Raison*, **14**, 47–74.

Manuscript received 22nd January, 1968.

RESPONSE DECREMENT AS A FUNCTION OF EFFORT

BY

GLORIA FISCHER, WAYNE VINEY, JAMES KNIGHT and NEIL JOHNSON
From Washington State University and Colorado State University

Albino rats were given acquisition training on a 6 ft. enclosed straight alley runway which could be adjusted for inclinations of 0°, 20°, and 40°. After receiving equal training on each of the three effort conditions, three groups of subjects were given massed extinction trials for 30 min. on only one effort condition. Results indicated fewer responses during extinction under the greater effort conditions. These results support a relation between effort and response decrement in extinction, consistent with the predictions of Hull's theory of reactive inhibition.

INTRODUCTION

The inhibition (I_R) theory of extinction maintains that each response in an extinction situation generates some inhibition to its recurrence. The theory maintains further that increasing degrees of effort with the consequent increasing amount of inhibition should result in a faster rate of extinction. Typical tests of this theory (e.g. Mowrer and Jones, 1943; Applezweig, 1951; Maatsch, Adelman, and Denny, 1954) have been contradictory and inconclusive. Part of the lack of agreement in results stems from certain procedural problems connected with differential bar loadings in the Skinner box (Trotter, 1956; Kimble, 1961). Kimble states:

> . . . two procedural difficulties arise: (1) In order to train the animal to press a heavy bar it is usually necessary to begin by teaching it to press a lighter bar first. Such training, in effect, may give the animals extinguished on the light bar more training than those extinguished on a heavier one. (2) The problem of partial responses is sometimes critical. Such responses, which consist of bar depressions of insufficient magnitude to operate the food release mechanism, occur more frequently with a heavy bar than a light one. During training, the occurrence of these responses introduces uncontrolled schedule effects; and during extinction, one must decide whether to count them. (Kimble, 1961, p. 285.)

Such partial responses or subthreshold responses to the heavy loadings present no more of a problem than superthreshold responses to the lighter loadings. In both cases, effort is produced which is either not measured or inaccurately measured.

One approach to the problem of differential training is to give each subject an equal amount of training on all conditions under which extinction can take place, a procedure followed by Capehart, Viney, and Hulicka (1958) and by Quartermain (1964). In both studies rats were given equal amounts of training on 5, 40, and 70 gm. counterweighted bars in a Skinner box. In spite of this procedural similarity, Capehart *et al.* found decreasing resistance to extinction as a function of effort, whereas Quartermain failed to find such a relationship. (Other procedural differences in the two studies, to be discussed later, may account for these apparently contradictory findings.) One problem with both studies is that subjects were given all initial training trials on the lighter effort conditions, while final training trials were given on heavier effort conditions. Thus, while frequency of response on each effort condition during acquisition was carefully controlled, there may have been biases introduced because of recency effects and initial training effects. In addition, both studies failed to measure or account for the critical problem of subthreshold and superthreshold responses.

The elimination of under and/or over responding, along with elimination of biases introduced because of recency effects and initial training effects can probably not be accomplished by utilizing differential bar loadings in a Skinner box. Such problems might be minimized, however, by an alternative technique such as traversing straight alleys that could be inclined to require various amounts of effort. Such a procedure has the additional advantage of assuring increased homogeneity of response from one effort condition to another. Heterogeneity of response is a serious problem in testing effort in a Skinner box. For example, rats required to press a light bar typically do so with one paw, whereas rats required to press a heavy bar use a different set of muscles involving both paws or sometimes the whole body.

The present study, then, attempted to overcome some of the important procedural problems, e.g. under and over responding and heterogeneity of response, that have beset previous studies relating effort and extinction.

METHOD

Subjects and apparatus

Subjects were 18 male and 18 female albino rats 70 to 90 days old at the beginning of the experiment. The apparatus was a modified Graham-Gagne type straight alley runway (Woodworth and Schlosberg, 1954). The runway (72 × 5 × 10 in.) was hinged to a 10 × 4·5 × 10 in. startbox at one end and a 10 × 6 × 10 in. goal-box at the other end. The hinging arrangement made it possible to elevate the goal box to create runway inclines of any desired magnitude. Guillotine doors constructed of opaque plexiglass separated the startbox from the entrance to the alleyway and the goal-box from the end of the alleyway. The runway floor was overlaid with ¼ in. hardware cloth to prevent slipping on steeper inclines. The startbox, goal-box, and alleyway were covered by wood framed ¼ in. hardware cloth. The entire apparatus was painted a flat mid-grey.

Acquisition procedure

Subjects were gentled for a few mintues on each day following weaning so that prior to introduction to the runway they were comparatively tame. The first training session was preceded by 24 hr. of food deprivation. During subsequent training sessions subjects were maintained on a 22 hr. deprivation schedule. These 10 min. sessions were given in a darkened room (light level in the runway was less than 1 ft.-cd.). Training continued until subjects completed 60 acquisition trials consisting of 20 trials on each of three work conditions or inclines (0°, 20°, and 40°). These conditions were changed from trial to trial on a random basis with the exception that no single effort condition occurred on more than three successive trials.

After a subject was placed in the startbox, the guillotine doors separating the startbox from the runway and the goal-box from the runway were raised. After a subject left the startbox, the first guillotine door was lowered to prevent retracing. Also, following entry into the goal-box the second guillotine door was lowered. Inter-trial interval (timed from the closing of the second guillotine door) was 60 sec. with the exception of the first few trials in which a subject was allowed additional time to explore and consume the food reward. Reward consisted of small pellets of wet mash. Following each training session subjects were returned to their individual home cages where they were permitted to eat Purina Chow *ad lib* for a period of 1 hr. With the exception of the first few training trials, no partial responses were observed throughout the experiment.

Extinction procedure

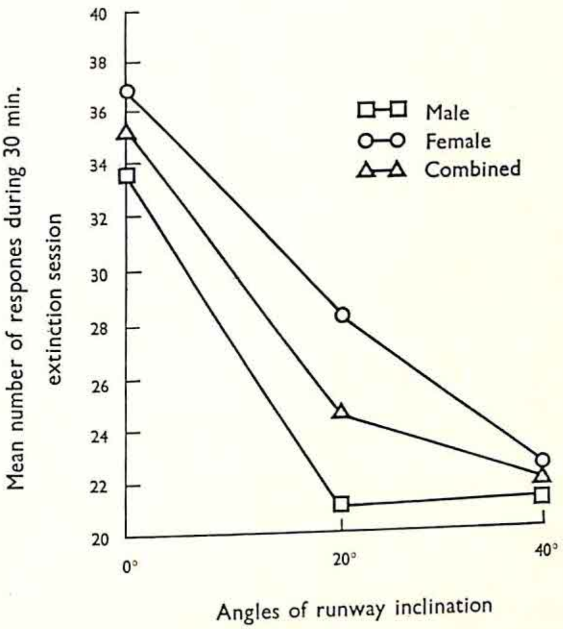
Following acquisition training subjects were assigned to one of three extinction conditions consisting of 0°, 20°, or 40°. Assignment to extinction condition was based on sex and mean response latency during acquisition. Approximately 24 hr. following the final acquisition session, subjects were tested for number of runway responses during a 30 min. extinction period. Inter-trial interval was 30 sec.

RESULTS AND DISCUSSION

Separate analyses of variance were performed on (1) average response latency (time from raising start box door to goal box entry) on later acquisition training trials and (2) the number of responses during the extinction session. The first analysis was performed in order to rule out the possibility that magnitude of runway inclination during acquisition affected running speed, which in turn, could have affected the number of responses made during extinction. Accordingly, comparisons were made for each rat's mean latency over the last three acquisition trials on each runway incline. The statistical design was a 3 × 2 × 3 factorial with repeated measures on acquisition incline (0°, 20° and 40°). The additional factors represented separation of subjects on the bases of sex and assignment to extinction incline (0°, 20°, or 40°). Sex was the only significant effect ($F = 4·94$, $d.f. = 1/30$, $p < 0·05$), with females running faster than males. Effects relating to both acquisition incline and assignment to extinction incline were insignificant ($F < 1·00$ for each factor). Interaction effects were also insignificant. Thus runway inclines, taken separately, did not produce differential running speeds during acquisition. In addition, group running speeds were equivalent prior to the extinction session, i.e., assignment of subjects to extinction runway incline was indeed random with respect to running speed.

The second analysis, a 3×2 factorial design, compared the number of responses during the 30 min. extinction session as a function of angle of runway incline (0°, 20°, or 40°) and sex. Results revealed reliable differences in the number of responses during extinction between angles of runway inclination ($F = 20{\cdot}78$, *d.f.* 2/30, $p < 0{\cdot}001$). In addition there was a sex difference ($F = 5{\cdot}04$, *d.f.* 1/30, $p < 0{\cdot}05$). The nature of the differences resulting from effort and sex are shown in Figure 1. As can be seen from the figure, female subjects were more resistant to extinction than male subjects. This finding is consistent with the fact that females ran faster than males in acquisition training. For both sexes responses to extinction clearly declined as a function of effort. In contrast to the highly significant difference between 0° and 20° (for 0° vs. 20° $F = 19{\cdot}6$, *d.f.* 1/30), however, the difference between 20° and 40° inclinations was not reliable ($F < 1{\cdot}00$).

FIGURE 1



Number of runway transversals during extinction as a function of sex and runway inclination.

The finding of less resistance to extinction after training in a runway requiring greater effort is clearly consistent with a reactive inhibition interpretation. However, resistance to extinction was decreased by extinction in an alley inclined 20°, but not decreased significantly further by training in an alley inclined 40°. Thus, inhibition from effort in this situation may not accrue in a linear manner. Instead, it may accrue in a non-linear manner.

In the main, these results are supportive of other studies (e.g. Capehart *et al.*, 1958; Mowrer and Jones, 1943) that found decreased resistance to extinction as a function of effort, and in contrast with studies failing to find such a function (e.g. Lawrence and Festinger, 1962; Quartermain, 1964). The apparent contradiction deserves close examination. Lawrence and Festinger (1962), utilizing a straight alley runway inclined to 25° and 50°, found that rats showed increased resistance to extinction on the 50° runway. This finding was interpreted as supportive of dissonance theory. An important difference between Lawrence and Festinger's procedure and that of the present study is that their subjects were trained only on the specific effort condition upon which extinction took place, whereas our subjects were given equal training on all effort conditions. Another important difference is that Lawrence and Festinger utilized spaced extinction trials (five spaced trials per day to a criterion of non responding) whereas the present study utilized massed extinction trials on a single day. I_R would be expected to dissipate substantially during spaced extinction trials, possibly attenuating an otherwise observable effort effect.

Further, there should be greater dissipation of I_R following the 50° condition, hence more recovery when a subject is put back in the extinction situation. In any case, most pertinent to effort and extinction are data based on massed trials of sufficient number for the effects of effort (presumably I_R) to accumulate.

The other study yielding results opposed to those of the present study and to those of the Capehart *et al.* study, was one by Quartermain (1964). One possible reason is that his subjects were comparatively overtrained (i.e. 360 acquisition trials, as opposed to 60 acquisition trials in the present study and 90 in the Capehart *et al.* study). Larger amounts of training during acquisition may produce sufficient physical conditioning to the larger effort requirements so as to adapt out the effect of effort during extinction. This possibility is presently being explored. Another important difference may have been the age and weight of subjects in the three studies. This may be especially critical in accounting for Quartermain's failure to replicate the Capehart *et al.* study, since he used 200 to 300 day old rats (mean body wt. 220·5 gm.), whereas Capehart *et al.* used rats which were closer to 100 days in age (mean body wt. approximately 160 gm.). A 70 gm. weight undoubtedly represents a far greater effort condition for a 160 gm. rat than for a 220 gm. rat, especially when the latter has received four times as much conditioning to the weight as the former. Thus, because of weight and training conditions, the Quartermain study used far lighter effort requirements than the study by Capehart *et al.* Correspondingly less effect of effort on extinction would, therefore, be expected.

The present experiment indicates that when equal training is given on variable effort conditions (including control for initial training experience and recency effects), and when subthreshold and superthreshold responses are eliminated, resistance to extinction is decreased by effort. In accord, however, with the suggestion offered by Lewis (1965), further investigation of the relation between extinction and effort is needed in a wider variety of situations.

References

Applezweig, M. H. (1951). Response potential as a function of effort. *J. comp. physiol. Psychol.*, **44**, 225–35.

Capehart, J., Viney, W., and Hulicka, I. M. (1958). The effect of effort upon extinction. *J. comp. physiol. Psychol.*, **51**, 505–7.

Kimble, G. A. (1961). *Hilgard and Marquis' Conditioning and Learning*. New York: Appleton-Century-Crofts.

Lawrence, D. H., and Festinger, L. (1962). *Deterrents and Reinforcements: The Psychology of Insufficient Reward.* Stanford: Stanford University Press.

Lewis, M. (1965). Psychological effect of effort. *Psychol. Bull.*, **64**, 183–90.

Maatsch, J. L., Adelman, H. M., and Denny, M. R. (1954). Effort and resistance to extinction of the bar-pressing response. *J. comp. physiol. Psychol.*, **47**, 45–50.

Mowrer, O. H., and Jones, H. M. (1943). Extinction and behaviour variability as functions of effortfulness of task. *J. exp. Psychol.*, **33**, 369–86.

Quartermain, D. (1965). Effect of effort on resistance to extinction of a bar-pressing response. *Quart. J. exp. Psychol.*, **17**, 63–64.

Trotter, J. R. (1956). The physical properties of bar-pressing behaviour and the problem of reactive inhibition. *Quart. J. exp. Psychol.*, **8**, 97–106.

Woodworth, R. S., and Scholsberg, H. (1954). *Experimental Psychology*. New York: Holt, Rinehart & Winston.

Manuscript received 21st January, 1968.

AUDITORY SIGNAL DETECTABILITY AS A FUNCTION OF PRE-EXPERIMENTAL SHOCK

BY

H. A. KUECHLER and P. C. DODWELL

From Ontario Hospital, and Queen's University, Kingston

A neurotic, schizophrenic and normal control group detected a 1,000 cps. 2 sec. tone with an *a priori* probability of occurrence of 0·5 against a background of continuous white noise during a specified 4 sec. interval. An index of signal detectability, d′, and a decision criterion, β, were calculated from the YES–NO responses for 150 control trials, and 150 trials which were preceded by free electric shock at maximum tolerance level. The results show a decrement in signals detected for the neurotic and schizophrenic group as compared to the normal control, whose detection rate improved significantly. The decision criterion remained unaffected.

INTRODUCTION

The theory of signal detection (TSD) (Swets, Tanner and Birdsall, 1961) suggests that the measure of signal detectability, d′, should only be affected by changes of the physical parameters of the signal and noise distribution, but not by observer variables which may be motivational or attentional. The decision criterion, β, is expected to change as costs and values of the decision outcome are altered or the *a priori* probability of signal occurrence is changed.

Swets (1961) demonstrated changes in the decision criterion as the result of changes in monetary rewards and losses in the direction predicted by TSD. However, Zwislocki, Maire, Feldman and Rubin (1958) did not find any significant changes in the detectability of auditory signals through the introduction of monetary rewards. The effect on the decision criterion is not discussed. Swets and Sewal (1963) similarly obtained no significant changes in the detectability of signals when verbal incentives and monetary rewards were introduced.

The present experiment examines the effect of non-response-contingent pre-experimental electric shock on the auditory signal detection performance of normal, neurotic and schizophrenic groups of subjects. If the shock serves to increase arousal, one might expect an improvement in performance (detectability) for normal subjects. The non-contingency of the stress, however, should leave the decision criterion in the TSD framework unaffected.

Specific hypotheses for the three groups were that normals would increase their detection performance under the shock condition, but that neurotics would show the type of decrement or perceptual avoidance behaviour under heightened stress proposed by Inglis (1961), and that the schizophrenics would have overall a lower detection rate which would show either a decrement or no change under stress. The latter hypothesis follows from the work of Stilson and Kopell (1964), who have shown that schizophrenics have considerable difficulty in detecting visual signals against a background of noise, and Malmo, Shagass and Smith (1950) provide evidence that schizophrenics are behaviourally inert to thermal stress but have above-normal autonomic responsivity. However, no clear prediction can be made about schizophrenics from this experiment, since behavioural and auditory functions respond differently. What is measured in the detection tasks is behaviour, although Swets (1961) emphasizes the importance of factors in the central nervous system in the analysis of detection studies.

Subjects

The two patient groups of eight subjects each were selected from the female admission ward of the Ontario Hospital. These two groups were selected on the basis of their conference diagnosis which reflects the consensus of three psychiatrists and ancillary clinical staff. The case history provided information to ensure the absence of psychotic or organic symptoms for members of the neurotic group, and similarly, only schizophrenics were chosen who had a history of hallucinatory experiences and absence of organic symptoms. The absence of visual and hearing defects was determined from the physical examination on admission or annual check-up for all subjects. Eight normal subjects were unpaid volunteers from the hospital staff.

Apparatus

A 1,000 cps. tone from a Heathkit Audiogenerator was presented, together with taped white noise, through fluid cushioned KOSS PR 4 earphones. The noise level was continuous and constant for all groups. The signal was presented randomly on 75 out of 150 trials. Onset and duration of a trial for 4 sec. was marked by illumination of a 15 watt light bulb, followed by an intertrial interval of 10 sec. The intertrial interval and signal duration were controlled by Hunter timers. A signal could occur variably within the 4 sec. trial interval for 2 sec., and this was occasioned by manual closure of a telegraph key which started the timer. No closure of the key produced no signal for that trial and constituted a blank, or "noise," presentation.

The subject recorded his decision by checking in a YES or NO column on a prepared sheet after each trial.

Procedure

The subjects of all three groups received 150 practice trials after their threshold for the signal tone against the white noise background had been determined, and the detection rate adjusted to 68–75 per cent. They were informed about the *a priori* signal probability of 0·5, but were given no knowledge of results. One hundred and fifty experimental trials followed the practice trials for half the subjects in each of the three groups. The other half was told that as part of the experiment pain tolerance threshold for electric shocks would be determined. The electrodes were 0·5 cm. in diameter, 2 cm. apart, and placed on the left forearm for right-handed subjects and on the right forearm for left-handed subjects. The shock intensity was increased in steps from zero to some value which was found to be barely tolerable. This was done by changes in voltage while the current remained constant for all subjects at 200 mA. Three 0·3 sec. shocks were given after this within the next 30 sec. The 150 experimental trials followed immediately. A further 150 trials were given during the same day which constituted the non-stressed condition while those subjects who received this condition first had the stressed or shock condition later. The intersession interval ranged between 45 min. and 2 hr.

RESULTS

The YES–NO responses of the subjects after each trial indicate his decision on whether a tone signal was presented or not. From the total number of correct signal identifications ($P_{SN}(A)$), and the proportion of affirmative responses to blanks ($P_{N}(A)$), or False Alarms, the decision criterion β, and index of signal detectability d′, were calculated from tables of the normal curve (Edwards, 1960). The values obtained agree closely with those given by Freeman (1964).

Table I shows the values of d′ and β for the normal, neurotic and schizophrenic groups under both standard and stressed conditions. An analysis of variance on d′ which takes into account the fact that the two stress conditions were measured on the same subjects (Winer, 1962, p. 302), indicates that, overall, d′ is not different in the three groups. However, the interaction of groups × stress condition is reliable ($p < 0\cdot 01$, $F = 13\cdot 67$; 2,21 *d.f.*) and Table I shows that it is due to an increase in d′ in the normal group, but decrements in the two patient groups. This was confirmed by analysis of the simple effect for each group separately (Winer, 1962, p. 232). The increase in d′ for the normal group is reliable ($p < 0\cdot 01$, $F = 11\cdot 00$; 1,21 *d.f.*) as is the decrease for the neurotic and schizophrenic groups ($p < 0\cdot 01$, $F = 9\cdot 67$ and $p < 0\cdot 05$, $F = 6\cdot 50$ respectively).

Inspection of the β values in Table I suggests that the schizophrenics adopted a more lax criterion than the other groups, and that all groups adopted a stricter criterion under stress. Analysis of variance of the β values showed that there is a significant difference between groups overall, attributable almost entirely to the schizophrenic group ($p < 0\cdot 05$, $F = 3\cdot 78$; 2,21 *d.f.*). However, the decrease in β values under stress which occurred in each group is not reliable overall ($p < 0\cdot 1$, $F = 2\cdot 45$; 1,21 *d.f.*), nor for any one group alone. This could be a real, but small effect, and might bear further investigation, especially since the data suggest a bigger change in the schizophrenics than in the others.

Essentially the same conclusion is reached by analyses of covariance of the stressed d′ and β values, with corresponding d′ and β values in the unstressed condition as covariates.

DISCUSSION

The results show that the detectability of auditory signals against a background of white noise is significantly improved in normals when compared with a decrement in patients diagnosed as neurotic or schizophrenic when free electric shock is administered

TABLE I

INDICES OF SIGNAL DETECTABILITY AND DECISION CRITERIA VALUES FOR STANDARD AND PRE-EXPERIMENTAL SHOCK CONDITIONS

Group	*Experimental condition*			
	Standard		*Stressed*	
	d′	β	d′	β
Normal	0·98	1·136	1·27	0·815
	1·71	0·732	2·49	0·532
	1·05	1·093	1·37	0·916
	0·65	1·252	1·25	1·588
	0·14	1·031	1·01	0·999
	1·56	1·283	1·74	1·136
	1·66	1·316	1·69	1·175
	0·78	0·808	1·14	0·695
Mean	1·066	1·081	1·495	0·982
Neurotic	0·77	0·755	0·50	0·831
	1·33	0·669	1·36	0·751
	0·87	1·313	0·39	1·112
	1·66	1·596	0·54	1·209
	1·09	1·378	0·80	1·192
	0·58	1·085	0·16	1·017
	1·11	0·874	0·61	0·989
	1·94	1·168	1·95	0·768
Mean	1·169	1·105	0·789	0·984
Schizophrenic	1·37	3·601	0·20	1·259
	0·57	2·010	0·06	1·058
	1·82	0·342	1·45	1·605
	0·80	1·155	0·80	0·864
	0·62	1·780	0·50	1·159
	0·20	1·197	0·18	1·071
	0·24	1·214	0·07	1·085
	1·34	2·390	1·22	2·620
Mean	0·870	1·711	0·560	1·340

before the experimental session. Thus our principle hypotheses are confirmed. The theory of signal detection does not predict changes in signal detectability except as a result of changes in the signal-to-noise ratio or other variations in the physical characteristics of the stimulus. Response contingent aversive stimuli should affect the decision criterion β. Since β showed no reliable change in this experiment one may assume that the non-contingent electric shocks preceding the experiment had at most a quite small effect on the observer's decision criterion. However, the finding of a reliably different decision criterion for schizophrenics suggests that a new look at perceptual abnormalities in schizophrenics is required.

The finding of increased detectability in the shock condition for normals appears to conflict with Horn and Venables' finding of poorer discrimination between pairs of clicks when weak shocks are applied (Horn and Venables, 1964). However, in their experiment shock was administered in conjunction with each trial, so the results of the two experiments are not really comparable. Symons (1963) found that a heteromodal stimulus (such as olfactory, proprioceptive, auditory) enhanced the detection of a visual stimulus in 83 per cent. of his subjects, but the effect only occurred if the heteromodal stimulus did not attract attention to itself. One could argue that in the Horn and Venables experiment electric shock acted as a distractor which resulted in poorer performance on the auditory task.

The result of the normal group's performance may be interpreted to support an arousal theory (Duffy, 1957) in the sense that prior stimulation with electric shocks had an activating effect and led to greater discriminability of signals. Fuster (1958) enhanced the discrimination of tachistoscopically displayed stimuli by monkeys through stimulation of the nuclei of the mid-brain reticular formation, the system most probably involved in arousal or activation phenomena.

The decrement in the detection performance of neurotics and to some degree of the schizophrenic subjects might be explained on the basis suggested by Inglis (1961, p. 264). Neurotic introverts are thought to be autonomically more labile or over-reactive than normals under stress, and under a sufficiently high level of stress perceptual avoidance behaviour is instituted, which could result in a lowered detection performance such as was found in this experiment. Malmo *et al.* (1950) found their chronic schizophrenics to be more aroused physiologically through the application of painful thermal stimuli than normals. In this respect they may be similar to neurotics, thus accounting for a process analogous to perceptual avoidance behaviour. However, no introversion-extraversion data were obtained from the present sample.

Whether the shock-induced stress operates peripherally to tighten musculature for better reception, or centrally as would be postulated by proponents of arousal theory is not answered by these results, but might be investigated with the use of muscle relaxants which have no CNS depressing effect, and amphetamines which have a central effect. Experiments by Mackworth (1965) suggest that amphetamines may not have any other effect on the detectability of signals than to prevent a decline in the vigilance of the subject.

The authors are indebted to Dr. J. S. Pratten, Superintendent, Ontario Hospital, Kingston, whose interest and support made this study possible.

REFERENCES

DUFFY, E. (1957). The psychological significance of the concept of "arousal" or "activation." *Psychol. Rev.*, **64**, 265–76.

EDWARDS, A. L. (1960). *Experimental Design in Psychological Research*. New York: Holt, Rinehart and Winston.

FREEMAN, P. R. (1964). *Tables of d′ and β*. Med. Res. Coun., Applied Psychol. Res. Unit, Cambridge. Mimeograph.

FUSTER, J. M. (1958). Effects of stimulation of brain stem on tachistoscopic perception. *Science*, **127**, 150.

HORN, G., and VENABLES, P. H. (1964). The effect of somesthetic and acoustic stimuli on the threshold of fusion of paired light flashed in human subjects. *Quart. J. exp. Psychol.*, **16**, 289–96.

INGLIS, J. (1961). In EYSENCK, H. J. (Ed.), *Handbook of Abnormal Psychology*. New York: Basic Books.

MACKWORTH, J. (1965). The effect of amphetamine on the detectability of signals in a vigilance task. *Canad. J. Psychol.*, **19**, 104–10.

MALMO, R. B., SHAGASS, C., and SMITH, A. A. (1950). Responsiveness in chronic schizophrenics. *J. Personal.*, **19**, 359–75.

STILSON, D. W., and KOPELL, B. S. (1964). The recognition of visual signals in the presence of visual noise by psychiatric patients. *J. Nerv. Ment. Dis.*, **139**, 209–21.

SWETS, J. A. (1961). Is there a sensory threshold? *Science*, **134**, 168–77.

SWETS, J. A., and SEWAL, S. (1963). Invariance of signal detectability over stages of practice and levels of motivation. *J. exp. Psychol.*, **66**, 120–26.

SWETS, J. A., TANNER, W. P., and BIRDSALL, T. G. (1961). Decision processes in perception. *Psychol. Rev.*, **63**, 301–40.

SYMONS, J. R. (1963). The effect of various heteromodal stimuli on visual sensitivity. *Quart. J. exp Psychol.*, **11**, 211–20.

WINER, B. J. (1962). *Statistical Principles in Experimental Design*. Toronto: McGraw-Hill.

ZWISLOCKI, J., MAIRE, F., FELDMAN, A. S., and RUBIN, H. (1958). On the effect of practice and motivation on the threshold of audibility. *J. acoust. Soc. Amer.*, **30**, 254–62.

Manuscript received 12th December, 1967.

EFFECTS OF GROUPING ON THE PUPILLARY RESPONSE IN A SHORT-TERM MEMORY TASK

BY

DANIEL KAHNEMAN,* LINDA ONUSKA and RUTH E. WOLMAN

From Harvard University Center for Cognitive Studies

Subjects heard strings of nine digits for immediate recall, either at a monotone 1/sec. rate or in groups of three digits separated by pauses. Concurrent measurements of pupil size show a steady dilation in the monotone condition, and brief dilation-constriction waves during the pauses of grouped presentation. The results are consistent with hypotheses concerning rehearsal in the two modes of presentation. Pupillary dilations apparently accompany episodes of covert rehearsal.

Introduction

When a string of digits is presented for immediate recall at a steady rate, and the subject's report is also paced at that rate, systematic changes occur in the size of his pupil (Kahneman and Beatty, 1966). During the loading phase, the subject's pupil dilates with every digit that he hears; during the unloading phase, his pupil constricts with every digit that he reports. In the original article, the steady increase of pupil size during the loading phase was ascribed to a steady increase of effort involved in rehearsing the digits. The present study was designed to test the relation between covert rehearsal and pupillary dilations.

Ryan (1967) has found that the introduction of pauses between successive groups of three digits markedly improves the recall of nine-digit strings. She suggests that such a grouped presentation imposes a coherent organization of rehearsal activity; rehearsal occurs mainly during the pauses, and is often restricted to the immediately preceding group. On the other hand, rehearsal is typically cumulative and repetitive in the ungrouped monotone presentation: the subject repeatedly attempts to rehearse the entire string, starting at the beginning (Murray, 1967; Reynolds and Houston, 1964; Ryan, 1967). If indeed pupil changes indicate rehearsal, the rhythm of presentation, which controls rehearsal, should be reflected in the time-course of pupillary dilations and constrictions.

Method

The subjects were paid volunteers, 14 from Harvard University and two from Radcliffe College. Photographs of the subject's right eye were made at 2 per sec. on infrared film with a modified Kodak K-100 camera, from which the shutter had been removed, and which was driven by an external motor. Tape recorded pulses activated a General Radio Strobolume, filtered for infrared with a Kodak 87C filter. The method has been described in greater detail elsewhere (Kahneman and Beatty, 1967).

Twenty random strings of nine digits, no digit repeated, were presented by tape recorder. The digits were synchronized with the pulses, ensuring co-ordination between the pictures of the pupil and the sequence of stimuli. In the ungrouped presentation, digits were recorded to correspond with pictures 5, 7, 9 . . . 21, at a monotone rate of 1 per sec. In the grouped presentation, digits corresponded to pictures 5, 6, 7 . . . 12, 13, 14, . . . 19, 20, 21.

The subject was instructed to withhold his response until after picture 25, then report the digits verbally, at his own pace. Several practice trials were given before the first pictures were taken. The first experimental trial belonged to the grouped condition. The two modes of presentation were alternated on successive trials.

Results

The difficulties that subjects experience with recall of nine digits forced adoption of a lenient method of scoring. A response was considered correct if it included a single omission or transposition of a pair of digits. With this method of scoring, the mean number of correct trials (out of a possible 10) was 7·38 for the grouped, 4·25 for the ungrouped presentation; the difference between the two modes of presentation is highly significant ($t = 5{\cdot}35$, 15 *d.f.*, $p < 0{\cdot}001$). This result confirms earlier reports of the effect of pauses (Mackworth, 1965; Ryan, 1967) and of grouping in rehearsal (Wickelgren, 1964, 1967).

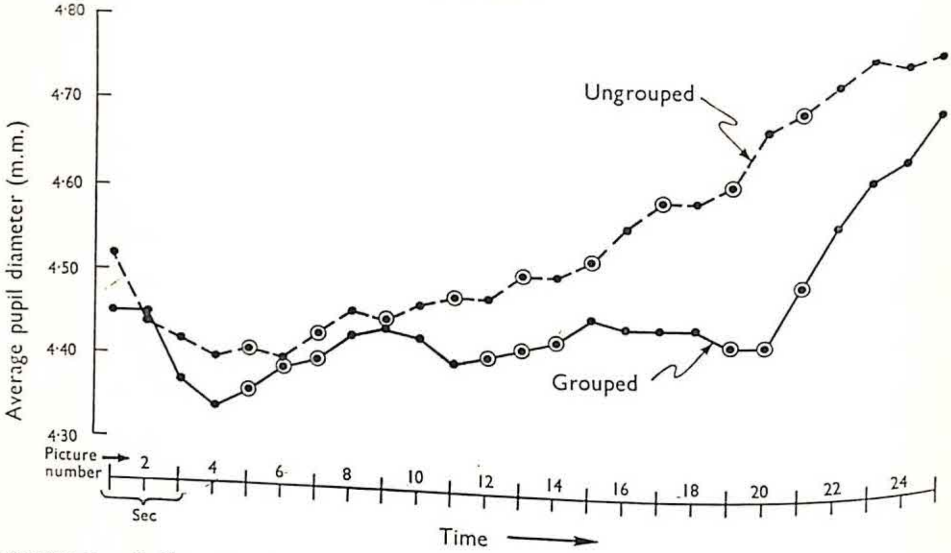
---

* Now at Department of Psychology, Hebrew University, Jerusalem, Israel.

An average pupillary curve was obtained for each subject under each condition, considering only those trials on which his response had been correct. Incorrect trials were rejected to avoid contamination by emotional responses accompanying failure. Four subjects were not included in the analysis of pupillary data because they had less than three correct trials in the ungrouped presentation. Analyses of pupil data are therefore based on 12 subjects.

The mean of the individual curves are given in Figure 1. As in the earlier study

FIGURE 1



Average pupil diameter for 12 subjects during grouped and ungrouped digit presentation. Digits were heard at the circled points.

(Kahneman and Beatty, 1966), pupil size in the ungrouped condition increases as a nearly linear function of the number of items already presented. In the grouped condition, pupil size increases very little during the presentation of the digits (pictures 5–21). The difference between the two conditions is significant ($t = 2{\cdot}69$, 11 *d.f.*, $p < 0{\cdot}025$). On the other hand, the steep dilation that occurs in pictures 21–25 is larger for the grouped than for the ungrouped presentation ($t = 2{\cdot}40$, $p < 0{\cdot}05$).

Of particular interest is the evidence for scalloping of the curve in grouped presentation. The pupil dilates slightly at the beginning of both pauses and constricts slightly as the pauses end and new digits are heard. The summed changes of pupil size from pictures 10 to 12 and from 17 to 19 are significantly negative ($t = 2{\cdot}53$, $p < 0{\cdot}02$, one-tailed). It is worthy of note that comparable constrictions do not occur anywhere else in either of the two curves.

Discussion

The pupillary results are in close agreement with Ryan's analysis of rehearsal activity. The cumulative nature of rehearsal in the ungrouped condition should require progressively more effort as digits are presented. A steady increase of pupil size during the presentation of the string confirms that prediction. On the other hand, rehearsal during grouped presentations is assumed to be intermittent and non-repetitive. Indeed, several of our subjects reported that they rehearsed only the second group of digits during the second pause. The results suggest that the subject engages in brief and undemanding episodes of rehearsal at the beginning of each pause, then relaxes in anticipation of the next group of digits to be heard. In this view of subjects' activity during grouped presentation, a relatively large effort might be required to pull together the three groups of digits at the termination of the presentation. The relatively steep dilation which is observed on pictures 21–25 supports that expectation. There is no reason to believe that the difference between the grouped and the ungrouped curves at that time is due to a ceiling effect; the relation between baseline and change in pupil data is minimal within the range of pupil size obtained in the present study (Kahneman and Beatty, 1967; Kahneman, Beatty and Pollack, 1967).

The absence of a pupillary loading function with the grouped presentation is not an isolated finding. In another experiment (Kahneman and Peavler, 1968), subjects heard a series of eight common nouns at a rate of 1 item per 4 sec., under free recall instructions. The presentation of each item was followed by a brief wave of dilation-constriction, without any sustained increase of pupil size during the course of the task. Cumulative rehearsal was probably minimal in that situation as well as in the present experiment.

The present experiment may be viewed as a typical exercise in construct validation (Cronbach and Meehl, 1955), which brings together a tentative theory and a tentative measure of rehearsal. The confirmation of predictions supports the validity of both theory and measure. It should be emphasized that precisely paced tasks and time-locked averaging of the pupil response are essential to this type of analysis, particularly when it is applied to autonomic responses which are consistent in direction but small in size.

This study was supported by National Institute of Mental Health grant PO1-MH-12623, and National Science Foundation grant GS-1153, to Harvard University, Center for Cognitive Studies. We are indebted to Joanna Ryan for very helpful discussions.

REFERENCES

Cronbach, L. J., and Meehl, P. E. (1955). Construct validity in psychological tests. *Psychol. Bull.*, **52**, 281–300.

Kahneman, D., and Beatty, J. (1966). Pupil diameter and load on memory. *Science*, **154**, 1583–5.

Kahneman, D., and Beatty, J. (1967). Pupillary responses in a pitch-discrimination task. *Percept. Psychophys.*, **2**, 101–5.

Kahneman, D., Beatty, J., and Pollack, I. (1967). Perceptual deficit during a mental task. *Science*, **157**, 218–9.

Karneman, D., and Peavler, W. S. (1968). Incentive effects and pupillary changes in association learning. *J. exp. Psychol.*, in press.

Mackworth, J. F. (1965). Presentation rate repetition and organization in auditory short-term memory. *Canad. J. Psychol.*, **19**, 304–15.

Murray, D. J. (1967). Overt *versus* covert rehearsal in short-term memory. *Psychon. Sci.*, **7**, 363–4.

Reynolds, J. H., and Houston, J. P. (1964). Rehearsal strategies and the primacy effect in serial learning. *Psychon. Sci.*, **1**, 279–80.

Ryan, J. F. (1967). *Grouping and Short-Term Memory*. Ph.D. dissertation, Cambridge University.

Wickelgren, W. A. (1964). Size of rehearsal group and short-term memory. *J. exp. Psychol.*, **68**, 413–9.

Wickelgren, W. A. (1967). Rehearsal grouping and hierarchical organization of serial position cues in short-term memory. *Quart. J. exp. Psychol.*, **19**, 97–102.

Manuscript received 10th November, 1967.

BOOK REVIEWS

Freud and his Early Circle: The Struggles of Psycho-Analysis. By Vincent Brome. London: W. Heinemann Ltd. 1967. Pp. xii + 275. 63s.

This is a story of the calamities and quarrels of the pioneers of psycho-analysis. It is a story that has often been told before, but Mr. Brome brings to his task some new information, including unpublished correspondence as well as published material that has become available since the appearance of the Jones biography. Although his sympathies are clearly with Freud, Mr. Brome manages to view the many disagreements, personal no less than professional, with praiseworthy detachment. For example, he is more than fair to Adler and does much to rehabilitate Ferenczi in his last, unhappy days. There is an Appendix on Havelock Ellis, based on a considerable correspondence with Freud now apparently in the author's possession. This is a tempting morsel which one may hope Mr. Brome will see fit to expand into a genuine re-appraisal of this strange, vain and now almost forgotten man.

O. L. ZANGWILL.

Fundamentals of Measurement (Second Edition). By N. M. Downie. New York: Oxford University Press. 1967. Pp. xiii + 482. 52s.

The full title of this book is *Fundamentals of Measurement: Techniques and Practices.* Intending readers who fail to take note of the sub-title may be disappointed, for it is an unusually essential statement about the chosen content. Handbooks on psychometrics may be expected to contain an extensive discussion of the principles of measurement, levels of measurement, the concept of error, probability, statistical methods of analysis, etc. In the case of Professor Downie's work such topics are limited to a rather light-weight coverage in the first quarter of the book, and thereafter he is almost entirely concerned with the description of tests. Although this description entails some critical evaluation, the first impression on reading it is that much of the material would be more suitably left for discovery in the various test agency catalogues than brought together as the core of a textbook. But this judgement is premature and too severe. Professor Downie is writing primarily for the American classroom teacher and school counsellor, workers whose contact with psychological measurement does not, presumably, require extensive theoretical knowledge. As a relatively compact statement for busy people of what tests are available for what purposes, and which are worth using, the book succeeds. However, it is difficult to see how it can also be a fully satisfactory text for the audience of students of psychology and education which is implied in the author's Preface. They, one suspects, would need rather more of the "Fundamentals" and less of the "Practices."

Parts II ("Achievement Tests") and III ("Measurement of Intelligence, Special Abilities, Adjustment, and Interests and Attitudes") contain some interesting points of comparison for the British reader. The idea of the teacher as a skilled writer of multiple-choice test items, the account of the keeping of 'anecdotal records' ("Several training sessions should be set up, in which each (school) staff member brings in at least one anecdote to be evaluated by the group."), and the enthusiasm for student evaluation of University teachers are far removed from the traditional conception of the teacher's rôle in this country. On the other hand, the effect of 'external testing programmes' on the school curriculum, the claims pro and con objective and essay tests, and the recognition of the value of small University classes strike a very familiar chord. For the most part, the arguments in Parts II and III are clearly stated, though there are occasions when the desire for brevity has unfortunate results. The five-page presentation of "theories of intelligence" is an example. Here, discussion of the concept of a group factor is virtually limited to a pictured shaded common area of two of three overlapping ellipses, Vernon's hierarchical structure of human abilities is described in terms of "top" and "bottom" without the benefit of a diagram (and with mis-numbered references), and the impression is conveyed that Thurstone's multiple factor theory is in some way opposed to the idea of group factors. This over-condensation is in marked contrast to, say, the leisurely detail with which the evaluation of interests is treated.

The book is in its second edition, and there are signs of haste in the revisions made. Thus, although on p. 70 there is a reference to the latest edition (1960) of the Stanford-Binet intelligence scale which points out that measurement is now in terms of standard score deviation IQs, an observation which is repeated on p. 272, p. 165 bears the following illustration of a "worst-answer" multiple-choice achievement test item:—

5. The following are examples of standard scores.
A. Wechsler IQs. +C. Stanford-Binet IQs.
B. Army General Classification Test Scores. D. T-scores.
E. stanines.

(+ = the "correct" answer)

And the inconsistency is emphasized on p. 275 where the section on the Stanford-Binet, which earlier stated "Binet IQs are currently a type of standard score," is drawn towards its conclusion with the suggestion that many people feel the scale's IQ scores should be replaced by standard scores.

Haste seems to be reflected too in the scattering of proof-reading errors, which are sometimes genuinely misleading. For example (italics are the reviewer's): "An objective test is one in which *the same results* are produced *regardless of who scores it*. If errors are made, they are brought about by carelessness and inaccuracy on the part of *the scorer, not the examinee*. On subjective tests results emerge that, in many cases, reflect the emphases, values, and idiosyncracies *of the scorer*." (pp. 14–15). This statement is either meaningless or bedevilled by the reversal of "scorer" and "examinee" in the second sentence. Or, when comparing essay and objective tests:— "The assessment of such processes is not confined to the peculiar domain of the essay test provided that the *essay test builder* is skilled in the art of test item writing." (p. 209, para. 7). Again meaningless, unless we read "objective test builder" for "essay test builder."

In short, the book has its uses but it is hardly to be preferred to the more adequate treatments of psychological testing by Anastasi or Cronbach.

D. J. BRUCE.

Psychological Measurement and Prediction. By Paul Horst. Belmont, California: Wadsworth. London: Prentice-Hall, 1967. Pp. vii + 455. 72s.

This is a difficult book to assess. An initial cursory examination gives the impression of a not entirely happy attempt at a taxonomy of test procedures and applications, leading into a systematic mathematically-oriented treatment of the basic concepts of descriptive statistics. The book is offered for use "both as an advanced undergraduate text and as a reference work for teachers, research workers, and others who are not specialists in psychological-measurement theory." As so often happens with dual or multi-purpose texts, the author's discourse emerges as not always on the same level, and the result strikes one as somewhat incongruous.

This reviewer, consequently, started reading with some degree of misgiving, but this, after the four introductory chapters had been disposed of, was fairly quickly dispelled. Dr. Horst's exposition, in the main part of the work, is both clear and stimulating—at times one might say even aphoristic. This is especially so in the chapter on Factor Analysis, for which the author presents a more than usually convincing case. The core of the book, however, is to be found in Chapters 11–16, in which, starting from the statistics of a single item, the author proceeds to discuss test statistics as functions of item-statistics, and thence to maximum value statistics, effects of chance, test homogeneity and test score and error. Later chapters deal with problems of reliability, specificity, validity, prediction and item analysis. Nearly every chapter concludes with mathematical proofs, in both scalar and matrix notation, of the principal formulae presented in the text.

Two main limitations of the work are: First, although the introductory material calls attention to other "domains" of measurement (e.g. sensory, motor, personality), little attention is paid to domains other than the cognitive. Accordingly, for the undergraduate, at least, who may be concerned more with the experimental than, say, the educational field, the relevance of much of the discussion to his interests may not always be obvious. Second, it is perhaps a pity that the scope of the book had not been extended to cover sampling and inferential statistics. Tests of significance receive little more than passing mention: indeed the concept of significance, as commonly understood, along with that of confidence levels, is something for which Dr. Horst has little use. He prefers the cost-utility concept of decision making of Cronbach and Gleser, since, as he puts it, "the problem is to estimate relative gains and losses for alternative actions, and to weigh them against the probability of success and failure on the basis of the experimental results." This is probably fair, particularly in relation to what Horst calls "real life decision making situations", but the continued importance of understanding the conventional approach seems indisputable, and to have had the relevant statistics included in Dr. Horst's scheme would have been both interesting and valuable.

BORIS SEMEONOFF.

Attention and Performance. Edited by A. F. Sanders. Amsterdam: North-Holland Publishing Co. 1967. Pp. xii + 452. £6.

Here is a valuable survey of current research in several important areas of human psychology, being the proceedings of a symposium held at Driebergen in August, 1966. The topics covered in the course of 46 papers by 59 authors include reaction times, vigilance tasks and short-term memory, looked at in many different ways. The most important treatments of these topics involve either the empirical study of physiological correlates or theoretical notions of information processing.

The book is divided into six sections entitled Single-channel theory and information processing, Reaction processes (these two sections occupy the first half of the book), Physiological correlates of attention and reaction time, Short-term memory and information processing, Eye movements and visual search, and Long-term performance. Studies of reaction times and notions of information processing are clearly dominant over the rest; maybe this represents the present emphasis in research. Each section is introduced by a general review paper and since many of the subsequent contributions tend to be surveys of research programs rather than definitive research reports, there is a high review capacity. I have found this book a valuable source of information about what is currently happening in studies of human performance, and it has also proved useful in teaching students at an advanced level. But it would have been much more useful for teaching (at the moment there is a great need of a text book here) if it had been possible so to choose the contributors to the symposium as to achieve a wider coverage and a better balance in reporting research on human attention and performance.

But the ultimate function of a review such as this must be to outline and evaluate current trends in research as they are represented in this book—in so far as this can be done in a short space. Many of the conceptualizations of information processing stem from the original theory of the psychological refractory period of Welford, who contributes a powerful and comprehensive review to the first section. Subsequent papers apply Welford's ideas to a variety of related experiments, modifying the theory in small ways, but, more important, showing too uncritical an acceptance of the central idea. The single-channel hypothesis has become established with an apparent authority greater than its empirical foundation justifies; it has become a conceptual framework for the description of reaction time experiments in which the signals are presented in pairs, separated by a short interval of time. Essentially, the hypothesis says that the reaction to each signal of the pair is an independent process and the two processes cannot overlap in time—so the second signal is delayed if it arrives during the reaction to the first. Accordingly for a suitable interval between the two signals the reaction time to the second signal will have a variance about twice that of the reaction time to the first; but so fundamental a property of the model appears not to have been investigated—not even after fourteen years.

Lack of critique is also illustrated by the several different uses of the term "information." Most contributors use this word in ways that are derived, more or less, from Communication Theory; this is the most common usage in Psychology. But some of the contributors, notably Wolfendale, Pickett, Sanders and Ter Linden, use the term in a sense related to statistical theory; more exactly, these authors are thinking in terms of a model based on the sequential probability ratio test. Now "information," whether in Communication Theory or in statistics, is an exactly defined concept. Several different definitions have been proposed, all of them related mathematically, and Information Theory, meaning the mathematical properties of these information measures, is a precise and powerful technique in the construction of theoretical arguments. The same precision should attach to the meaning of "information" in Psychology. But here, as in most psychological writing, the term is used with a variety of ill-defined shades of meaning and with scant regard for the original mathematical precision; the conceptual power of Information Theory is thereby wasted.

The measurement of cortical evoked potentials in reaction time experiments is a recent and very interesting development. Whereas the information-theoretic study seeks to relate merely the input and output variables of the experiment, this technique looks at the reaction process at an intermediate stage, recording the intervening gross neural activity. The technique is difficult and involves computer averaging over hundreds of trials. This in turn leads to difficulties of interpretation: for example, some components of the cortical evoked response are presumably related to the stimulus input, others to the production of the response. These latter will vary in time with the duration of the reaction process—but it is not at present possible to obtain meaningful records of individual reactions. A somewhat similar evaluation applies to the recording of eye movements during visual search. Such measurements give additional information about the operation of control processes which serve to select information for perceptual processsing.

The sections on short-term memory and long-term performance (principally vigilance experiments) are small relative to the wide spectrum of current research in these fields. The papers included in the symposium give a sample rather than a cross-section of current ideas and trends, and it is not possible to say much by way of outline or criticism without going beyond the material in the present book. There is a penetrating review by Posner of current concepts and ideas in the study of short-term memory and a theoretical treatment by Sperling of his results on the processing of the visual image. For vigilance tasks Jerison presents a critical analysis of the uses and shortcomings of activation theory and Taylor discusses the analysis of vigilance data by means of signal detectability theory.

What began as a symposium to facilitate the exchange of results and ideas in the study of human performance has resulted in an interesting and valuable book. It is more up-to-date and presents a wider spectrum of ideas than any textbook, however advanced. But its present usefulness will quickly fade, because it does not have the coherence of treatment or the clarity in conceptual exposition that gives a textbook continuing usefulness. Of course, it is not, and was never intended to be, a textbook; I am making the comparison because the present absence of any detailed text means that for a short time this book will be pressed into an unintended use as an up-to-date source book for advanced students. In two or three years it will have no greater value than a typical journal volume of similar age (it is published as an additional volume of *Acta Psychologica*). But by that time I hope there will be another symposium covering much the same ground and serving much the same purpose, such is the present usefulness and value of this volume of proceedings.

D. R. J. LAMING.

Sensory Inhibition. By Georg von Békésy. London: Oxford University Press; Princeton University Press. 1968. Pp. x + 265. £4 1s.

This book comprises the Langfeld Memorial Lectures delivered by von Békésy at Princeton in 1965. It displays that same lone virtuosity evident in his earlier work as collated by Wever in *Experiments in Hearing* and the overall effect is very similar: that of a collection of brilliant experiments on the same theme, but not intimately related, described in detail but not given the continuity and exposition normally associated with a monograph. Who but a true scientist in the off-beat Galton tradition would have an orchestra play in an enclosed courtyard and a marquee to demonstrate the inhibitory funneling of the auditory system? Again, von Békésy holds out for the superiority of certain psychophysical measurements over physical measurements, an unconventional view, but in its limited context quite correct. Only the initiate is likely to appreciate these qualities, and the new student should in fact be discouraged from reading the book, because it is also true in this type of creative scientist that the work of other authors tends to be neglected.

All this makes the book unsuitable for regular information retrieval on the topic of sensory inhibition; but that topic is already fairly well served by Ratliff and not all books have to be intermediate level texts. There is room for occasional collectors pieces, which is what this volume is, in content, format and price.

MARK P. HAGGARD.

Elements of Psychophysics, Volume 1. By Gustav Fechner. Translated by Helmut E. Adler. New York: Holt, Rinehart & Winston Inc. 1966. Pp. xxxi + 286. 56s.

It is perhaps hardly surprising that Fechner's *Elemente der Psychophysik* had never previously been translated into English. His primary aim was to bolster rationally the status of the psychical in a mid-nineteenth century world which, as he saw it, was half drugged by scientific materialism and half intoxicated with thoughtless romance. Psycho-physics was to him a means to an end; what was intrinsically interesting in it to others was soon culled and publicized notably by Wundt, G. E. Müller and, later, directly into English by Titchener. The text itself, faithfully rendered as it has been by Dr. Adler, is frankly still as nearly unreadable in English as it is in all the syntactic contortions of the original German. Earlier translators more readily expended their labour and skill on other, headier, examples of his work such as *Nana oder Über das Seelenleben der Pflanzen*.

The present edition is therefore perhaps a labour of love, or a work of piety, and any who still recognize and respect such acts will applaud their instigators—in this instance Professor Boring and Dr. Davis Howes. It seems doubtful whether anything fresh that is essential to Fechner's scientific achievement is likely to be established by the wider scrutiny a translation will allow. And so far as his attitude, approach and personal preoccupations are concerned, more is revealed in Volume II. But at least it can be said that what has been done has been done well. Dr. Howes in his introduction claims that Dr. Adler's " . . . English version is readable, and yet faithful to the original." The first part of the claim is technically, the second unreservedly, acceptable. No more, and no less, could have been expected of a translation made to celebrate the centenary of the original work.

Is the *Elemente* a "great work"? Or does it only happen to constitute a nodal point in the complex currents of mid-nineteenth century thinking? It does not, like Newton's *Principia*, present at one fell swoop a closely integrated and almost impregnable structure from which a vast number of deductions and explanations are possible. Nor, like the writings of say John Hughlings Jackson and Sigmund Freud, is it pregnant with a large variety of unpredictable offspring. Except from motives of historical or human curiosity there is no particular reason to turn back to it a century later. But through all its untidiness and alternate concern with issues, on the one hand metaphysical and sometimes even mystical, on the other doggedly and minutely empirical, it maintains a principle—namely that there are determinate, measurable relations between the events within the organism which are produced by physical stimulation. And it goes further and presents specific statistical procedures by which these relations may be defined even though we are largely ignorant of the chain of internal organic mechanisms involved. This principle and the procedures—or others largely derived more or less directly from them—remain quite central to psychology, and to other disciples such as economics, today. We should never forget that it was Fechner and no-one else who introduced them. In this sense the *Elemente*, if not a "great," is certainly a "seminal," work and merits translation into the English which has so largely usurped the place of Fechner's own language in scientific communication.

It would be a great pity if this enterprise were left uncompleted, but on the face of it the auspices do not seem good. In his preface Dr. Howes' last words are "The translation of the second volume remains an undertaking for the future," and perhaps more ominously (and not quite honestly) the binding carries no reference to the fact that it shelters only Volume I. Volume II contains much, including the fascinating chapter on imagery which, if less germane technically to contemporary psycho-physics, is essential to an understanding of the scope of Fechner's intentions. If translator, editors and publishers can be encouraged by an economically satisfactory reception of the part already done to finish the job, librarians ought with one accord to order this book.

R. C. OLDFIELD.

Human Information Processing. By Lawrence J. Fogel. London: Prentice-Hall. 1968. Pp. ix + 420. 70s.

A book with this title might be expected to deal with such important topics as the processing of information in immediate memory, categorization and absolute judgements, adaptation to visual and perceptual rearrangement, pattern recognition, detection and discrimination, information reduction in thinking and problem-solving, to name just a few. On these topics there is little or nothing in this book, which is a reprint of a work, *Biotechnology: Concepts and Applications*, first published in 1963. The new title is justified by the omission of large sections of the earlier work, but by nothing else. There has been no revision or re-editing, and the book starts at "Section B, page 61."

The main contents are chapters on the senses; a section on decision-making which contains a chapter on manual tracking and shorter discussions on a number of topics, such as the evolution of the brain; and a section which deals with the control of motor movement, the design of machine controls, and "techniques for sensing and interpreting information concerning the human operator without his overt co-operation." The book gives a competent account of these areas with less attention to method and more to mathematical development and description than is usual in psychology textbooks. It is out-of-date: of the 113 references to the chapter on vision none were published in the sixties. Colour vision is discussed with reference to the work of Land—the names of Young, Helmholtz and Hering are not mentioned. I. Pollock, C. W. Parky, K. J. W. Craig, O. B. Hebb and other hitherto unknown yet vaguely familiar names make a transient appearance in these pages.

This book may have some usefulness for engineers who need to know something about "the human operator," but any psychologist who buys it on the strength of the title and the publishers' reputation will undoubtedly be disappointed.

M. TREISMAN.

Vision and the Eye. By M. H. Pirenne. 2nd Edition. London: Chapman & Hall; New York: Barnes & Noble. 1967. Pp. xvi + 224. Hardback 50s. Paperback 25s.

This excellent book, first published 20 years ago, has now been republished in a second edition. It is perhaps the best available introduction to the physiology of the eye. The author modestly recommends it to beginners and non-specialists. Although the presentation is attractive enough to lure Laodiceans, there is plenty of solid information for research students and their seniors. It covers the physiology and structure of the mammalian eye and the insect eye, the response of the nervous system to light, absolute sensitivity and quantum fluctuations, visual acuity and colour vision. Pirenne sets these problems in a wider context by comparing the eye's performance with the natural limits on the sensitivity and accuracy of any possible physical device. He formulates these problems quantitatively, and carefully avoids subjective and introspective methods. This would rule out a large number of perceptual experiments, but he does not attempt to cover perception, apart from a new section on the perception of perspective in photographs; this is a book of physiology, not psychology.

The second edition has recognisably the same structure as the first, but it has been rewritten throughout to bring it into line with visual research since 1948. Pirenne observes that "there never was a moment, at which the chief part of what existed was not old" (Macaulay's epigram on the British Constitution, applied by Aubert to our knowledge of physiological optics). This is belied by Pirenne's reference list: more than half the works he cites (138 out of 252, or 54 per cent.) date from the last two decades, since the first edition came out. Physiology advances faster than politics.

New sections have been added on the role of perspective in the perception of paintings and photographs, taking into account the psychological fact that the observer is usually aware of the flat surface he is looking at: the directional properties of the retinal cones (Stiles-Crawford effect): and the statistical uncertainty of response in the single cell preparation from Limulus, including some unpublished material by Hartline. The section on colour vision is brought up to date with a summary of the recent studies of photosensitive pigments made by Marks, Dobelle and MacNichol (1964), Rushton (1965) and Brown and Wald (1964), which almost solve the colour matching problems raised 150 years ago by Thomas Young—and which support Young's own theory. Land is also mentioned. There are a few gaps. Barlow's and Rushton's important work on light- and dark-adaptation probably came out too recently to be included, but it is odd that the chapter on "the two eyes and the brain" contains no mention of Julesz. But overall the revision has been carried out thoroughly. The book should sell well until a third edition becomes due in 1988.

S. M. ANSTIS.

The Human Operator in Complex Systems. Edited by W. T. Singleton, R. S. Easterby and D. Whitfield. London: Taylor & Francis Ltd. Pp. xx + 198. 57s. 6d. Also published as Vol. 10, No. 2, of the journal *Ergonomics*.

Ergonomics from the psychologist's point of view during the 1940s was largely based on the design of displays and controls. However, those who went in for this kind of work quickly came to realise three points which caused them to widen their horizons. Firstly, a performance which is satisfactory in a short-term laboratory task, may not be so if it is continued over longer periods—in other words, the time-scale of an operation needs to be considered. Secondly, many subtle social factors which enter into real-life work situations can profoundly affect performance and need to be taken into account as an integral part of any ergonomic designing of a job. Thirdly, the psychologist called in to discuss ergonomic factors of particular pieces of industrial or service equipment often wondered whether the thinking behind the whole installation had been sufficiently radical: he came to realize that it is not enough to throw together several individually designed pieces of equipment, however excellent each might be. Rather, design of all the equipment needs to be considered together as a unified *system*, taking into consideration the purpose to be achieved, the optimum allocation of tasks between men and machines, appropriate training procedures and, behind all these, questions of cost—it would, for example, often be possible to design a machine which could replace a human operator completely, but quite uneconomic to do so.

This kind of work has in the past depended on, and contributed substantially to, fundamental knowledge about human capacity and performance. The present book aims at summarizing the stage that thinking in this field has reached and at pinpointing some of the areas in which further basic knowledge is required. After a long preface explaining the *systems approach* to work design, there are six sections dealing with (1) Systems design methods, (2) Systems-analysis techniques, (3) Allocation of function between man and machine. (4) Methods of describing tasks, (5) Training, job-aids and maintenance, and (6) Applications. The reader is advised to begin with the last three sections. Without the concrete examples contained in these as a background, the papers in the first three sections often appear trite, although this is far from true in fact—the simplicity and elegance of their ideas is the result of much long and hard thinking over many past years.

It would be impossible here to discuss all the 21 contributions individually, but special mention should be made of R. J. Beishon's excellent paper on the complex skill involved in process-control, of R. S. Easterby's clever marriage of theory and application in the design of displays, of R. Conrad's lucid exposition of the principles of designing postal codes, of A. Crawford's penetrating discussion of the social consequences of installing a computer in an industrial organization, and of the article by B. Shackel, D. Beevis and M. Anderson on the fundamental problems involved in a new meat-handling system for the London Docks. The last paper, by R. F. Chaillet, contains the Human Factors Requirements which are written into contracts placed with industry for the supply of equipment to the U.S. Army.

A. T. WELFORD.

Behavior. An Introduction to Comparative Psychology. By John B. Watson. New York: Holt, Rinehart and Winston. 1967. Pp. xxi + 439. $7.95.

This book was published first in 1914, and reappears now in a series of reprints of classics in experimental psychology. If one's knowledge of the history of modern psychology has come mainly from secondary sources, one's first reaction may be to welcome having a book like this forced upon one's attention. How well is this initial enthusiasm maintained?

To begin with, one certainly encounters the Watson one had been led to expect: the first chapter is a classic statement of behaviourism (a term whose original use seems to have been confined to animal psychology). Watson puts very clearly the behaviourist's dilemma. The immediate object of his attention is the behaviour of animals, and in this behaviour he begins to discern the kind of lawful relations of which a science is made up. But if psychology is to concern itself with mental life, then the importance of animal behaviour can lie (as Lloyd Morgan acknowledged) only in what it allows us to infer about the mental life of animals. Then, of course, the difficulty arises that we are denied the use of introspection, the only check on the correctness of our inferences. Orthodox psychology seemed to have constituted itself in such a way that the student of animal behaviour was permanently denied any important part in it. In such a situation the only answer is revolution, and Watson presents his ultimatum. "Either," he says, "psychology must change its viewpoint so as to take in facts of behavior, whether or not they have bearing upon the problems of 'consciousness', or labour will have to be withdrawn, and behavior must stand alone as a wholly separate and independent science." He continues uncompromisingly with a threat that the subsequent history of psychology has fulfilled: "Should human psychologists fail to look with favour upon our overtures and refuse to modify their position, the behaviorists will be driven to use human beings as subjects and to employ methods of investigation which are exactly comparable to those now employed in the animal work." The claim Watson is making now seems modest enough: that behaviour should be studied in its own right and not be constantly subject to "analogical interpretations." Having put the case for the scientific study of behaviour so clearly, Watson indulges briefly in the kind of optimism about its everyday usefulness—to "the educator, the physician, the jurist, and the businessman"—that came to play a much larger part in his later writing. He sees the psychologist as a behavioural engineer, a kind of consultant for the control of behaviour.

This chapter, together with some provocative speculations about the relations between language and thought (he predicts that a man without a larynx would be unable to think) stands very much by itself. The book is a curiously ill-combined mixture of manifesto and textbook. The larger part of it is a convenient summary of methods and findings in the study of animal behaviour. Much space is devoted to leisurely and meticulous descriptions of apparatus and procedure, and to detailed analysis of work on instinct, habit formation, and sensory processes—work carried out for the most part without any licence from a self-conscious "behaviourism." Watson's textbook reputation is mainly as a propogandist and it is perhaps useful to be reminded here that he was also a resourceful experimenter. His interest in sensory processes and the study of discrimination is particularly strong, and there are a number of places in which he anticipates what we are likely to think of as more recent developments. He deals for instance (p. 68) with the use of certain unlearned response tendencies in the study of discrimination, and outlines techniques very similar to those currently used for teaching animals errorless discriminations (p. 69).

Watson does not devote much space in this book to the theoretical analysis of learning. Classical conditioning is mentioned briefly as a technique for the study of sensory processes, and for an explanation of habit formation Watson relies, having rejected the Law of Effect as too mentalistic, on the principles of frequency and recency. Nevertheless, it is worth pointing out that most of the explanatory devices—proprioceptive stimuli, mediating responses, and so on—that were to serve learning theorists so well for another 40 years, are used here by Watson.

M. H. SHELDON.

The Psychology of Communication. By George A. Miller. London: Allen Lane the Penguin Press. 1968. Pp. vii + 197. 35s.

This immensely readable book contains seven essays, of which only one, the inevitable "Magical Number Seven," has previously appeared in a psychological journal. The first three cover recoding and information-processing, the fourth deals with psychical research, and the final three are concerned with computers and psycholinguistics. The ideas in this book are original and interesting, and of great significance in the development of psychology.

The three 1956 essays on human information-processing are notable for their emphasis on coding and recoding, a hitherto-neglected phenomenon in psychological experiments. The concept of a limited information-processing capacity is also expounded in some detail, although there is some confusion between a capacity of "bits" and one of "bits per second," which deserved clarification. The three more recent essays on computers and psycholinguistics make interesting reading, particularly the most entertaining discussion of the impact of computers. In my opinion, the only weak point of an otherwise excellent book is the essay on psychical research, which consists entirely of a critical review of C. D. Broad's *Lectures on Psychical Research*. Since the essay lacks the originality and significance of the other six, I think it would have been better omitted. Apart from this one reservation, I would recommend this book to anyone with any interest in psychology.

D. P. FARRINGTON.

PUBLICATIONS RECEIVED

BOOKS

Automation Theory and Learning Systems. Edited by D. J. Stewart. London and New York: Academic Press. 1967. Pp. xi + 215. 63s.

A Stereotaxic Atlas of the Brain of the Pigeon (*Columba Livia*). By Harvey J. Karten and William Hodos. London: Oxford University Press. Johns Hopkins Press. 1968. Pp. ix + 193. £9 10s.

Advances in Parasitology. Vol. 5. Edited by Ben Dawes. London and New York: Academic Press. 1967. Pp. xvi + 318. 80s.

The Concept of Development: An Issue in the Study of Human Behavior. Edited by Dale B. Harris. London: Oxford University Press. Minnesota University Press. 1968. Pp. x + 287. 64s. cloth. 24s. 6d. paper.

Personality and Arousal: A Psychophysiological Study of Psychiatric Disorder. By Gordon S. Claridge. Oxford: Pergamon. 1967. Pp. xviii + 274. 70s.

How to Find Out in Philosophy and Psychology. By D. H. Borchardt. Oxford: Pergamon. 1968. Pp. vii + 97. 20s. Hard cover. 12s. 6d. Flexi-cover.

Psychological Diagnosis in Clinical Practice with Applications in Medicine, Law, Education, Nursing and Social Work. By Benjamin Pope and Winfield H. Scott. London: Oxford University Press. 1968. Pp. xii + 339. 81s.

Work and the Nature of Man. By F. Herzberg. London: Staple Press. 1968. Pp. xx + 203. 25s.

Non-Human Thought. By Jacque Graven. London: Arlington Books. 1968. Pp. 202. 30s.

Information Theory of Choice-Reaction Times. By D. R. J. Laming. London and New York: Academic Press. 1968. Pp. ix + 172. 63s.

Behavior-Genetic Analysis. Edited by Jerry Hirsch. London: McGraw-Hill. 1967. Pp. xvii + 522. 100s. 6d. $12.50.

Pleasure and Pain: A Study in Philosophical Psychology. By J. L. Cowan. London: Macmillan. 1968. Pp. 174. 36s.

The World of Ted Serios: "Thoughtographic" Studies of an Extraordinary Mind. By Jule Eisenbud. London: Jonathan Cape. 1968. Pp. 367. 50s.

Personality and Assessment. By Walter Misch Mischel. London and New York: Wiley. 1968. Pp. xii + 365. 75s.

Psychopharmacology: Dimensions and Perspectives. Edited by C. R. B. Joyce. London: Tavistock. Mind and Medicine Monographs. 1968. Pp. ix + 430. 63s.

Task Interruption. By Annie van Bergen. Amsterdam: North Holland. 1968. Pp. viii + 342. 105s.

The Structural Basis of Word Association Behavior. By Howard R. Pollio. The Hague: Mouton & Co. 1968. Pp. 96. 15 Dutch Guilders.

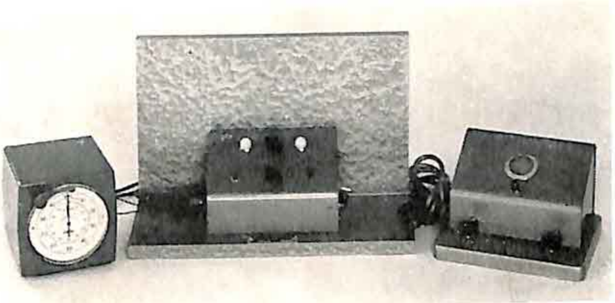
PAPERBACKS

POP 2 Papers: An Introduction to the POP 2 Programming Language; and Reference Manual. By R. M. Burstall, J. S. Collins and R. J. Popplestone. Edinburgh: Oliver & Boyd. 1968. Pp. 43. 21s.

L'enseignement programmé et les machines à enseigner en U.R.S.S. By A. W. Schestakow. Translated from the Russian by A. Kirchberger. Paris: Dunod. Sciences du Comportement. 1968. Pp. 218. 45,25 F.

Investigating Linguistic Acceptability. By Randolph Quirk and Jan Svartik. The Hague: Mouton & Co. 1968. Pp. 118. 14 Dutch Guilders.

Four Stimuli
Separable Consoles
Monitor Lights
"Get Ready" Signal



Measures Simple,
Discriminative,
and Choice
Reaction Times

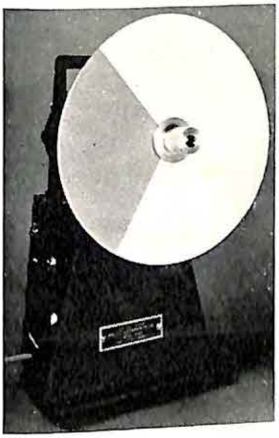
Model RT-5 Visual Choice Reaction Time Apparatus

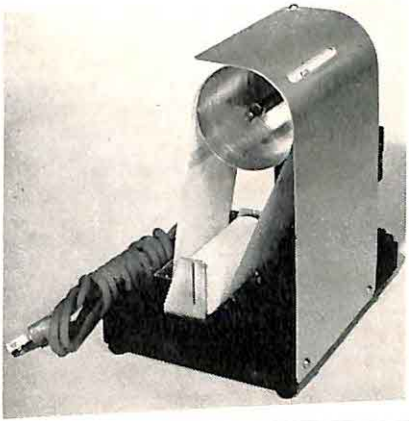
Harvard Tachistoscope

With solid state,
direct reading timer
Available in two and three field models

Colour Mixer

Heavy base
Quality Bodine motor
Easy to carry
Has series type speed control
Will handle discs up to 17″ in dia
A rugged unit suitable for student and research use





Memory Drum

Our model M1-A is of SIMPLE, DURABLE design
Has 3 exposure speeds, can be used with long or short word lists, is an ideal drum for use in class or laboratory

Electronic Voice Key

All solid state Model C-370

Complete catalogue available on request

Ralph Gerbrands Company

R G Co

ANNOUNCEMENT

PROPOSED EUROPEAN BRAIN AND BEHAVIOUR SOCIETY

We, the undersigned, wish to propose the formation of a EUROPEAN BRAIN AND BEHAVIOUR SOCIETY. The objects of the Society shall be the furtherance of scientific enquiry within those fields that bear on the interrelationships between brain and behaviour by holding periodic meetings at which papers are read and discussions held, by the dissemination of information and education materials made available as a consequence of research in the interrelationship of brain and behaviour, and by such other activities as may be decided upon by the Society.

A study group will meet in Rotterdam in the Spring of 1969 in order to continue our discussions in detail of the aims of the proposed Society and the direction in which it shall move, to consider and adopt its constitution, and to select its first membership. All persons interested in becoming a member of the Society are invited to write as soon as possible to Dr. A. Cowey (Institute of Experimental Psychology, South Parks Road, Oxford, England) stating their qualifications, experience, present work and interests. In addition, it would be appreciated if any opinions and suggestions about the general aims of the proposed Society could be expressed at the same time, and if applicants would state whether they are willing to deliver a paper at the first meeting of the Society, reporting scientific work that is considered to be in line with the aims of the proposed Society.

(*Signed*)

K. Akert, Zurich
A. Cowey, Oxford
M. Frankenhaeuser, Stockholm
H. G. J. M. Kuypers, Rotterdam
J. Paillard, Marseille
D. Ploog, Munich
J. Scherrer, Paris
L. Weiskrantz, Oxford

QUARTERLY JOURNAL
OF
EXPERIMENTAL PSYCHOLOGY

ANNOUNCEMENT

Publication Arrangements

With effect from 1st January, 1969 the Quarterly Journal of Experimental Psychology will be published for the Experimental Psychology Society by Academic Press Inc. (London) Limited. Subscribers will receive renewal notices from the Academic Press. The price of volume XXI will be £6. All transactions to take effect after 1st January, 1969 should be conducted with:

Academic Press Inc. (London) Limited,
Berkeley Square House,
Berkeley Square,
LONDON, W.1

or through booksellers as hitherto.

THE QUARTERLY JOURNAL OF
EXPERIMENTAL PSYCHOLOGY

Vol. XX 1968 Part IV

EFFECT OF INTERPOLATED EXTINCTION AFTER PARTIAL DELAY OF REWARD TRAINING ON SUBSEQUENT REACQUISITION AND EXTINCTION

BY

C. THOMAS SURRIDGE, KAREN R. MOCK,* and ABRAM AMSEL

From the Department of Psychology, University of Toronto

Four groups of albino rats were run four trials a day in a straight runway for 44 days. On the first 15 days, two groups were given continuous immediate reward (IR) and two groups a 50 per cent. schedule of 30-sec. partial delay of reward (PDR). On the next 15 days, one IR group and one PDR group were extinguished, while the other IR and PDR groups remained on their original schedules. In the third phase, all groups received 8 days of training on IR. Finally, all groups were given 6 days of extinction training. In the first extinction, PDR produced greater resistance to extinction than IR. In the second extinction period, the PDR group which had previously been given extinction and the two IR groups extinguished relatively rapidly and at approximately the same rates, while the PDR group which had not been extinguished was significantly more resistant to extinction than the other three groups.

INTRODUCTION

Increased resistance to extinction after partial delay of reward (PDR), relative to continuous immediate reward (IR), can be sustained through an extended period of suspended training and a block of IR trials interpolated between PDR acquisition and extinction (Donin, Surridge and Amsel, 1967). Essentially the same result was reported earlier for partial reward (PR) training (e.g. Ross, 1964; Theios, 1962). These results can be accounted for by assuming that greater response persistence is acquired under both PDR and PR, relative to IR conditions, and that this persistence depends on associative factors which can carry through suspended training, blocks of IR training, and changed motivation to affect behaviour in a later extinction phase.

Some recent three-phase experiments have contributed to our understanding of PR phenomena. Harris, Smith, and Weinstock (1962) found that extinction trials (Phase 2), interpolated between IR or PR training in Phase 1 and reacquisition under IR in Phase 3, resulted in an immediate shift toward higher asymptotic performance levels in Phase 3 than in Phase 1. However, a subsequent experiment of the same sort (Surridge, Boehnert and Amsel, 1966) indicated that if the interpolated extinction phase was extended very greatly, an "elation" effect was not found in Phase 3 IR reacquisition. The results of these two experiments suggested that, while a relatively brief extinction period might serve to raise the drive or arousal level of PR and IR subjects due to frustrative excitement, such motivational effects would be expected to disappear following a very extended extinction period (Surridge, Boehnert and Amsel, 1966, p. 565). This reasoning depends on some of the assumptions underlying

---

* Now at the Ontario Institute for Studies in Education.

frustration theory (Amsel, 1958, 1962). After some minimal number of rewarded trials, sufficient for the development of fractional anticipatory reward (r_R), nonreward in PR training results in primary frustration (R_F). R_F is the UCR for the conditioning of fractional anticipatory frustration (r_F), the antedating form of the aversive goal response to nonreward, which is assumed to have motivational properties. One indicant of such motivational properties is the Goodrich-type crossover effect—the initial inferiority and eventual superiority in acquisition of PR over IR performance (vigour), usually observed in the early segments of a response chain (Goodrich, 1959; Surridge *et al.*, 1966; Wagner, 1961).

Our argument may be summed up as follows: the effect of a few extinction trials would be to add conditioned frustrative excitement (r_F) to the motivational complex for both PR and IR subjects. Such an increase in drive would be reflected in higher asymptotic performance when these subjects were switched from extinction to IR reacquisition. On the other hand, since r_R and r_F are, by definition, classically conditioned responses, a very extended period of extinction should lead first to the extinction of r_R–s_R, and then to the extinction of r_F–s_F, the mechanism assumed to mediate persistence.

An implication of such reasoning was tested by Boehnert (1967) in a four-phase experiment. In the main experimental condition, PR training was followed by an extended extinction phase, and then by IR reacquisition and a second extinction phase (PR → EXT → IR → EXT). Extinction performance in the final experimental phase was very similar to extinction after IR training alone. On the other hand, omission of the extended, second-phase extinction period (PR → IR → EXT) produced the usual partial reinforcement effect (PRE), similar to the findings of Theios (1962).

It was proposed earlier that PDR and PR involve the same mechanisms; that, after the anticipation of immediate reward has developed in PDR training, delayed reward acts frustratively like nonreward. If this is the case, the earlier findings for PR (Boehnert, 1967; Surridge *et al.*, 1966) should also obtain for PDR. The experiment we report tests this proposition.

METHOD

Subjects

The subjects were 32 male Wistar albino rats, approximately 70–80 days old, supplied by Woodlyn Farms, Guelph, Ontario. They were housed in individual cages and maintained on *ad libitum* water. The daily ration of food was 10 gm. of Purina lab chow.

Apparatus

The apparatus was a straight enclosed wooden runway consisting of a $13\frac{1}{2}$ in. start-box, painted flat gray, and a 5 ft. 3 in. alley, painted flat black. The start box was 3 in. wide and narrowed to 2 in. while the width of the runway was $2\frac{3}{4}$ in. The goal area comprised the last 14 in. of the alley, and was separated from the rest of the alley by an aluminium guillotine retrace door. The runway was covered with a plexiglass lid. Subjects were not handled at the start of a trial but were maintained in the experimental room in individual wooden boxes with sliding aluminium lids, which served as entry boxes to the runway. These boxes were 9 in. × 3 in. × 3 in. (inside dimensions). The start door was of plexiglass, and dropped when an electrical circuit was broken by the experimenter. In the end-wall of the goal-box was a $1\frac{3}{4}$-in. square opening, $\frac{7}{8}$ in. from the goal-box floor, which allowed access to the food cup. A Model D Gerbrands unit, activated by interruption of the final photobeam, dispensed the food pellets. A Hunter timer was used to allow delay of operation of the Gerbrands Food Dispenser.

Five 1 ft. running measures were provided, beginning with the opening of the start door and ending 3 in. from the end wall of the goal area. A microswitch on the start door and five photobeams controlled five Standard Electric timers which recorded to 0·01 sec.

Procedure

Experimental design. The rats were randomly assigned to one of four groups, an IR–IR group, an IR-Extinction (IR–E) group, a PDR–PDR group, and a PDR–Extinction (PDR–E) group. These group designations cover the first two phases of the experiment. The third and fourth phases were the same for all groups and were, respectively, IR and extinction.

Habituation. The animals were placed on a 10-gm./day deprivation schedule 14 days prior to the start of the experiment, during which time they were handled and tamed. On the 2 days prior to the beginning of experimental training, they were allowed to explore the runway in groups of three for approximately 3–4 min. No food was present in the runway on exploration days.

Experimental training. The individual entry-box was placed at the end of the start box, and the entry-box door was raised. The start door was dropped by a remote switch as soon as the rat oriented toward it. When he entered the goal-box the retrace door was lowered. The rat was removed from the goal-box immediately after eating on rewarded trials; on all extinction trials confinement duration was 60 sec. The subjects were tested in two squads of 16, four from each group in each squad. The order of running subjects on the four trials of a day was the same, but was randomized from day to day. The intertrial interval was between 20 and 25 min. All 16 subjects in a squad were returned to their home cages after the day's trials and given their daily food ration 30–40 min. later.

In both extinction sessions, the rat was removed from the runway if it took more than 30 sec. to traverse any segment of the alley. A time of 30 sec. was recorded for that segment and for each of the remaining ones on that trial. A record was kept of all extinction trials on which the subject retraced (turned completely around) in the runway.

Experimental phases. The experiment was run in four phases. In Phase 1 subjects received a total of 60 trials. There were four trials per day throughout the experiment. Reward was two 250-mgm. Noyes pellets (500 mgm.). Groups PDR–PDR and PDR–E were rewarded immediately (I) on 50 per cent. of the trials and were delayed (D) in the goal-box for 30 sec. prior to reward on the other 50 per cent. of the trials. On any day there were two D trials and two I trials, and the daily sequences used were as follows: DIDI, IDDI, IDID, and DIID. These sequences were randomly assigned to subjects within a day so that all subjects did not receive the same sequence on any day. Groups IR–IR and IR–E were immediately rewarded throughout Phase 1. In Phase 2, Groups IR–E and PDR–E were run 60 extinction trials. The reinforcement schedule for Groups IR–IR and PDR–PDR remained unchanged in this phase. In the third phase, all subjects were run 32 trials to IR. For those in Groups IR–E, PDR–PDR, and PDR–E but not for Group IR–IR, this represented a change in the reward schedule from Phase 2. In Phase 4, all subjects were run 24 extinction trials.

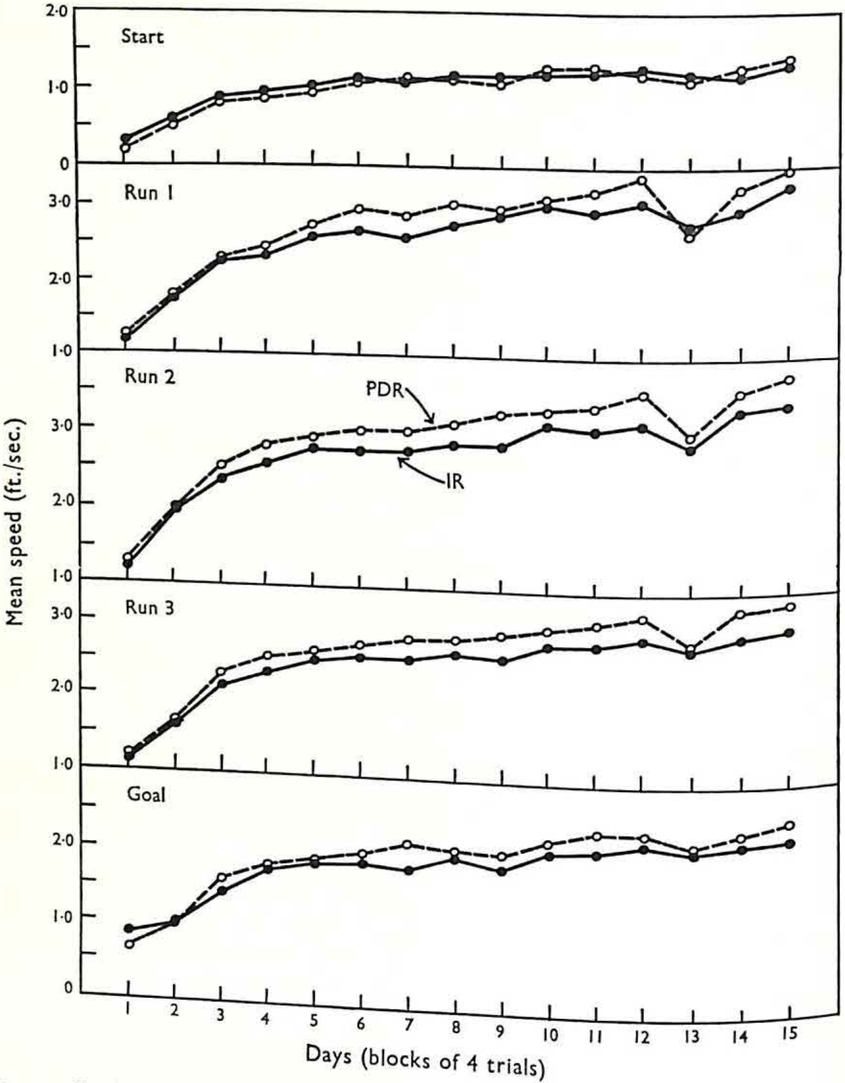
RESULTS AND DISCUSSION

Acquisition (Phase 1)

Analyses of variance were performed comparing the speeds of Groups IR–IR and IR–E, and of Groups PDR–PDR and PDR–E over the first phase for each of the five runway measures. As expected, neither the Group main effect nor the Group × Day interaction reached significance in any measure in either of the comparisons. Accordingly, the data in Figure 1 are for combined IR groups and combined PDR groups. Initial acquisition performance was very similar for IR and PDR, and remained so in the Start measure throughout acquisition, while average PDR speeds became greater in the three Run measures and in the Goal measure after the first few days of training. However, neither Group main effects nor Group × Day interactions suggested significant differences between IR and PDR in any measure.

The speed data for the last three experimental phases showed the same performance differences between the four groups in each of the five measures in each phase. The five measures were therefore combined to form a total alley measure, and converted to speeds (ft./sec.). The lower panels of Figure 2 show these overall speeds

FIGURE 1



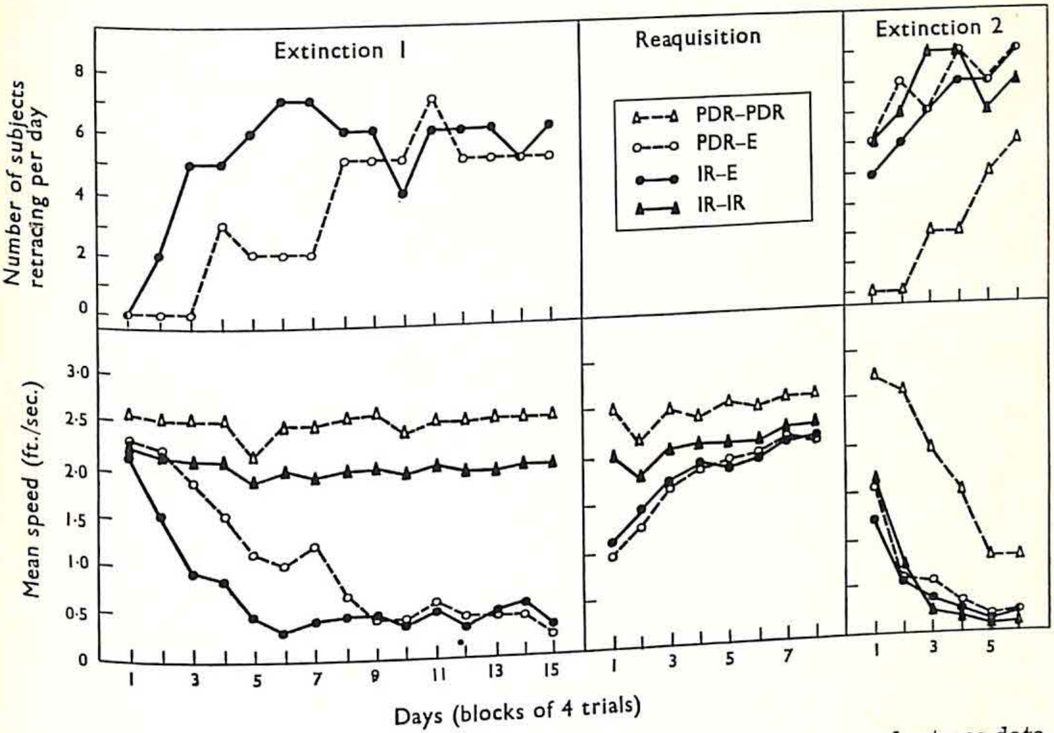
Mean speeds in Phase I for Groups IR–IR and IR–E combined and for Groups PDR–PDR and PDR–E combined in blocks of four trials.

for each group for the last three experimental phases; the upper panels present retrace data from the two extinction phases, plotted above the corresponding extinction speed data.

Extinction 1 (*Phase* 2)

Performance of Groups PDR–PDR and IR–IR, which were not extinguished in this phase, remained, of course, at approximately the same levels as in the acquisition phase. Both the speed and the retrace data reflect large differences in the initial extinction rates of Groups PDR–E and IR–E. This confirms earlier findings (Crum, Brown and Bitterman, 1951; Logan, Beier and Kincaid, 1956) to the effect that PDR leads to greater resistance to extinction than does IR. Over the last 7 days of the extinction phase, however, performance was essentially the same in both groups,

FIGURE 2



Mean total alley speeds plotted in blocks of four trials (lower panels), and retrace data plotted as the number of subjects retracing per day (upper panels) for the last three experimental phases.

whether in terms of retraces or speed. Analysis of variance on the speed data over Days 1–8 yielded a significant Group main effect, $F\ (1, 14) = 4{\cdot}71$, $p < 0{\cdot}05$, but not a significant Groups × Day interaction. Analysis over the last 7 extinction days indicated that neither the Group main effect nor the Group × Day interaction was significant ($F < 1{\cdot}00$).

Reacquisition (Phase 3)

Examination of the lower centre panel of Figure 2 reveals no tendency for Groups PDR–E and IR–E to show higher asymptotic speeds in the reacquisition phase than the nonextinguished groups. By the end of this phase they were performing at about the level of Group IR–IR. The group switched from PDR in the first two phases to IR in the third also shows no evidence of higher performance level than before. The reacquisition curves of Groups IR–E and PDR–E appear to be rather typical negatively accelerated acquisition curves. The results of this phase and the first two phases are essentially similar to those of Surridge *et al.* for PR (1966, Fig. 1, p. 567).

Extinction 2 (*Phase* 4)

Both the retrace and the speed data of the final experimental phase (right-hand panels of Fig. 2) indicate that Groups IR–IR, IR–E, and PDR–E extinguished very rapidly, and at about the same rate. Group PDR–PDR, however, extinguished more slowly than the other three groups, performance being similar to the initial extinction of Group PDR–E, and showing the initial positive acceleration characteristic of extinction following partially reinforced acquisition (Amsel, Rashotte and MacKinnon, 1966). This should be compared with the sharp initial drop and negative acceleration of the other three groups in extinction. Analysis of variance on the speed data comparing Groups PDR–PDR and PDR–E yielded a significant Group main effect, $F(1, 14) = 40{\cdot}10$, $p < 0{\cdot}01$, and a significant Groups × Day interaction, $F(5, 70) = 11{\cdot}33$, $p < 0{\cdot}01$, indicating that the two groups did in fact extinguish at different rates.

Analysis of the data of Groups PDR–E, IR–E, and IR–IR over the six extinction days revealed no significant Group main effect ($F < 1{\cdot}00$), though the Groups × Day interaction did reach a conventional level of significance $F(10, 105) = 2{\cdot}59$, $p < 0{\cdot}01$.

Donin *et al.* (1967) have shown that the partial delay of reinforcement extinction effect (the PDRE) will occur even when a period of suspended training and a block of IR trials are interpolated between PDR training and extinction. The results of the final extinction phase in the present experiment replicate this finding: the PDRE was clearly evident in the Phase 4 extinction comparison of Group PDR–PDR (PDR training followed by a block of interpolated IR trials) and Group IR–IR (continuously reinforced training throughout followed by extinction). However, the major findings of the present experiment are (*a*) that the persistence which develops in PDR acquisition and is sustained through IR can be eliminated by interpolating an extended extinction period between PDR and IR training, and (*b*) that when PDR is followed by extended extinction, there is no evidence of elated asymptotic effects in later IR reacquisition. These results are essentially the same as have been shown in PR training (Boehnert, 1967; Surridge *et al.*, 1966).

Our results support the proposition that similar, if not identical, mechanisms operate in acquisition and extinction under partial reward and partial delay of reward. In frustration theory terms, the introduction of a delay (when the subject has come to expect immediate reward) produces primary frustration. Later in training, the response-produced stimuli (s_F) from anticipatory frustration become counterconditioned to approach and lead to an extinction effect similar to the PRE. Moreover, the PDRE, like the PRE, can apparently be eliminated by extended extinction, implying, as we suggested earlier, the extinction of r_F.

It is important to note that our interpretation of PDR phenomena in terms of frustrative nonreward factors does not necessarily apply to continuous delay of reward (CDR); if subjects are never rewarded immediately, then delayed reward would seem to provide no basis for frustration in acquisition, and greater persistence should not develop in CDR. Earlier evidence regarding the effects of CDR on resistance to extinction has been equivocal (e.g. Cogan, 1966; Fehrer, 1956; Marx, McCoy and Tombaugh, 1965). Very recent evidence, however, indicates that if CDR acquisition is run out to an asymptotic level of performance (which has not been the case in earlier experiments), resistance to extinction is not very different to that for IR. Sgro, Dyal, and Anastasio (1967) found that asymptotic acquisition under continuous delays of 7·5 and 15 sec. produced resistance to extinction equal to that of an IR control group. Some unpublished data from our laboratory suggest that, by some criteria, asymptotic performance under CDR of 30 sec. actually produces *less* resistance to extinction than IR. The point we would make here is that, although similar mechanisms control acquisition and extinction under PDR and PR, those underlying CDR are not the same and will require a different theoretical analysis.

It is, of course, possible to interpret the present data and our other recent experimental findings in terms other than those of the conditioning model of persistence effects we have employed. For example, those who favour a discrimination hypothesis (e.g. Bitterman, Fedderson and Tyler, 1953) hold that the onset of extinction is more readily discriminable after IR training than after PR training; that PR subjects have difficulty discriminating the transition between acquisition and extinction because of their previous experience with runs of nonrewarded trials. In the present experiment, onset of the first extinction can be regarded as less discriminable following PDR than following IR if delay of reward is more similar to nonreward than is IR, accounting for slow extinction following PDR. The more rapid second extinction of the PDR group can be attributed to discriminative training between delay of reward and nonreward, in the first extinction session (Phase 2). Although this kind of interpretation appears to account for the PDR–E and IR–E data, there is evidence from the present experiment and elsewhere against the discrimination hypothesis as an adequate overall explanation of PR and PDR phenomena. Interpolating a number of IR trials between PR or PDR training and extinction would presumably serve to isolate PR or PDR from extinction and thus provide a sharp contrast in the reinforcing events at the onset of extinction, similar to that experienced after IR training alone. Several experiments, however, have shown that the relative persistence which results from PR or PDR acquisition can carry through suspended training, blocks of IR trials, changed apparatus, and even changed motivation and reinforcement (Boehnert, 1967; Donin, Surridge and Amsel, 1967; Ross, 1964; Theios, 1962). The extinction data of our Groups PDR–PDR and IR–IR represent a replication of some of these findings.

Among the other current interpretations of PR and PDR phenomena, Logan's interpretation (Logan, 1960; Logan, Beier and Kincaid, 1956) that subjects "stop r_G-ing" to cues following nonreinforcement is not ruled out providing it is reasonable to argue that PDR subjects do not make r_Gs during the 30-sec. period preceding reinforcement. Logan's theory states that "resistance to extinction is assumed to be increased by any training procedure which permits extinction of the fractional anticipatory goal response to the cues which occur on nonreinforced trials after the regular time of reinforcement" (Logan *et al.*, 1956, p. 70). Also, Capaldi's (1967) generalization decrement theory, in terms of the sequential effects of nonreinforcements (N–lengths) can specifically predict the results of our experiment. Moreover, Capaldi would predict (as would we) greater resistance to a second extinction if the first extinction was short, but less resistance if the first extinction was long (1967, pp. 117–124). On the other hand, cognitive dissonance theory as applied to PR and PDR phenomena (Lawrence and Festinger, 1962), would not seem to predict the effect on the second extinction of the Phase-2 extinction of Group PDR–E. Lawrence and Festinger explain greater resistance to extinction after PR and PDR training in terms of "extra attractions" that a subject develops for the experimental situation on nonreward or delay trials. Moreover, they state most explicitly that "These 'extra attractions' are not removed during extinction trials" (1962, p. 57). Presumably, their prediction as to the performance of Group PDR–E in the second extinction would be exactly opposite to our obtained results. The influential two-process analyser model of discrimination learning, which has also recently been offered as a theoretical contender in the explanation of the PRE (Sutherland, 1964, 1966), might also be applied to explaining our present results. If PDR is assumed, like PR, to increase the number of analysers switched into responding, and if the prolonged first extinction phase breaks the attachments between these analyser outputs and responding, the rapid extinction of the PDR–E group would not be unexpected. The question is then what this kind of attention theory says about the kind of condition represented by the PDR–PDR–IR–E group: Do the attachments between a number of analyser outputs and responding survive the interpolated IR training or not?

This work was supported by Research Grants GB-3772 from the National Science Foundation and APB-72 from the National Research Council of Canada.

REFERENCES

AMSEL, A. (1958). The role of frustrative nonreward in noncontinuous reward situations. *Psychol. Bull.*, **55**, 102–119.

AMSEL, A. (1962). Frustrative nonreward in partial reinforcement and discrimination learning: some recent history and a theoretical extension. *Psychol. Rev.*, **69**, 306–28.

AMSEL, A., RASHOTTE, M. E., and MACKINNON, J. R. (1966). Partial reinforcement effects within subject and between subjects. *Psychol. Monogr.*, **80**, (20, Whole No. 628).

BITTERMAN, M. E., FEDDERSON, W. E., and TYLER, D. W. (1953). Secondary reinforcement and discrimination hypothesis. *Amer. J. Psychol.*, **66**, 456–64.

BOEHNERT, J. B. (1967). *Investigation of the Within-Subject Extinction Effect Using a Separate-phase Method of Stimulus Presentation.* Unpublished Ph.D. thesis, Univ. of Toronto.

CAPALDI, E. J. (1967). A sequential hypothesis of instrumental learning. In SPENCE, K. W. and SPENCE, J. T. (Eds.). *The Psychology of Learning and Motivation. Vol. I* (*Advances in research and theory*). New York: Academic Press.

COGAN, D. C. (1966). Post-reinforcement delay in extinction: A failure to replicate. *Psychon. Sci.*, **6**, 343–4.

CRUM, J., BROWN, W. L., and BITTERMAN, M. E. (1951). The effect of partial and delayed reinforcement on resistance to extinction. *Amer. J. Psychol.*, **64**, 228–37.

DONIN, J. A., SURRIDGE, C. T., and AMSEL, A. (1967). Extinction following partial delay of reward with immediate continuous reward interpolated, at 24-hour intertrial intervals. *J. exp. Psychol.*, **74**, 50–3.

FEHRER, E. (1956). Effects of amount of reinforcement and of pre- and postreinforcement delays on learning and extinction. *J. exp. Psychol.*, **52**, 167–75.

GOODRICH, K. P. (1959). Performance in different segments of a response chain as a function of reinforcement schedule .*J. exp. Psychol.*, **57**, 57–63.

HARRIS, S. J., SMITH, G. S., and WEINSTOCK, S. (1962). Effects of nonreinforcement on subsequent reinforced running behaviour. *J. exp. Psychol.*, **64**, 388–92.

LAWRENCE, D. H., and FESTINGER, L. (1962). *Deterrents and Reinforcement.* Stanford: Stanford University Press.

LOGAN, F. A. (1960). *Incentive.* New Haven: Yale University Press.

LOGAN, F. A., BEIER, E. M., and KINCAID, W. D. (1956). Extinction following partial and varied reinforcement. *J. exp. Psychol.*, **52**, 65–70.

MARX, M. H., MCCOY, D. F., and TOMBAUGH, J. W. (1965). Resistance to extinction as a function of constant delay of reinforcement. *Psychon. Sci.*, **2**, 333–4.

ROSS, R. R. (1964). Positive and negative partial-reinforcement extinction effects carried through continuous reinforcement, changed motivation, and changed response. *J. exp. Psychol.*, **68**, 492–502.

SGRO, J. A., DYAL, J. A., and ANASTASIO, E. J. (1967). Effects of constant delay of reinforcement on acquisition asymptote and resistance to extinction. *J. exp. Psychol.*, **73**, 634–6.

SURRIDGE, C. T., BOEHNERT, J. B., and AMSEL, A. (1966). The effect of interpolated extinction on the reacquisition of partially- and continuously-rewarded responses. *J. exp. Psychol.*, **72**, 564–70.

SUTHERLAND, N. S. (1964). Visual discrimination in animals. *Brit. med. Bull.*, **20**, 54–9.

SUTHERLAND, N. S. (1966). Partial reinforcement and breadth of learning. *Quart. J. exp. Psychol.*, **18**, 289–301.

THEIOS, J. (1962). The partial reinforcement effect sustained through blocks of continuous reinforcement. *J. exp. Psychol.*, **64**, 1–6.

WAGNER, A. R. (1961). Effects of amount and percentage of reinforcement and number of acquisition trials on conditioning and extinction. *J. exp. Psychol.*, **62**, 234–42.

Manuscript received 8th May, 1968.

INFLUENCE OF ACCOMMODATION ON THE VIEWING OF AN OPTICAL ILLUSION

BY

MICHEL MILLODOT

From the Hunter Laboratory of Psychology, Brown University

This study demonstrates the role of microfluctuations of accommodation in the viewing of the concentric ring illusion. Four experiments were carried out. Experiment I consisted of changing the distance between the illusion and the eyes but retaining visual angle, luminance and pupil size constant. In Experiment II, the accommodation of the eyes is paralysed by a cycloplegic. In Experiment III, the illusion is shown to a monocular aphakic and in Experiment IV, the exposure time is varied between 1/100 and 5 sec. Furthermore, a series of observations compared the concentric ring illusion seen with both eyes simultaneously. The evidence provided by the results of these experiments plus others discussed provide evidence that the concentric ring illusion is caused by the microfluctuations of accommodation.

INTRODUCTION

Just over a century ago, Helmholtz (1867) described an optical illusion made up of black and white concentric circles (Fig. 1.) When this figure is observed, a propeller effect or a rotational sensation of radial lines is seen. This particular effect is due to the fact that along the brighter meridians, the black and white lines are sharp whereas in the other areas they are blurred and out of focus. Helmholtz attributed this phenomenon to the accommodation of the eye because he noted that as you change the distance between the pattern and your eyes other portions of the pattern become clear thus giving the impression that the clear radial lines are rotating. He also wrote "Dass die beschriebenen Erscheinungen von einer Asymmetrie des Auges herrühren is zunächst klar. . . ." The asymmetry of the eye to which he refers is nothing but astigmatism. Le Grand (1965) shares the opinion of Helmholtz but since Arnulf and Dupuy (1960) have shown that the microfluctuations of accommodation are astigmatic i.e. the ciliary muscle contracts asynchronously in different regions, he suggested that they were likely responsible for the propeller effect seen when looking at Figure 1. If the accommodative system of the eye is responsible, it is indeed appropriate to attribute the concentric ring illusion to the microfluctuations of accommodation since one notes the illusion even when the pattern and the head are fixed. The microfluctuations of accommodation first described by Arnulf, Dupuy and Flamant (1951) are of the order of 0·1 diopter in amplitude in photopic vision looking at a threshold target at optical infinity. Their dominant frequency is about 0–0·5 c/s. as was measured by Campbell, Westheimer and Robson (1959) using their infra-red optometer (Campbell and Robson, 1959). Moreover, Campbell and Robson (1958), pointed out that the time course of the movements of the optical illusion corresponds to that of the microfluctuations of accommodation.

Very recently, Evans and Marsden (1966) carried out an experiment in which they stabilized the retinal image of this illusion (among others). They used an after-image as their method of compensating for eye movements as this affords a perfect stabilization as compared to optical systems. They observed the concentric ring illusion before and after stabilization and found that the propeller effect disappeared when the image was stabilized. Thus, they concluded: "The absence of the radial propeller effect in stabilized vision seems to make it quite certain that the phenomenal lines are a function not of cortical scanning but of eye movements and consequent shifts of the image across the retina." The opinion that eye movements are responsible for this phenomenon is also given by Pritchard (1958) and Mackay (1957*a*) suggests that it may be of central rather than retinal origin. However, Evans and Marsden's method consists of placing an image of the very bright pattern on the retina and to observe its after-image. Consequently, they eliminate at the same time both the effect of the involuntary eye movements and of the accommodative system of the eye. Thus, it cannot be concluded that either eye movements or the microfluctuations of accommodations or an interaction of the two are responsible as this would demand that either one or the other factor be eliminated but not both at the same time. Moreover, it can be objected that observing the after-image of a pattern in order to note subtle variations within it may not be the most efficient way.

On the other hand, Mackay (1957*b*) stabilized the retinal image of this illusion. This was done by mounting a photographic reduction of the pattern on a stalk of 1·5 cm. long attached to a contact lens of + 60 diopters. He found that the concentric ring illusion was unaffected. Yet, with the use of this method eye movements are eliminated but the microfluctuations of accommodation are still present. From the results of the experiments by Evans and Marsden (1966) and by Mackay (1957*b*), it can be predicted that the accommodative system of the eye is responsible for the observation of the propeller effect seen when viewing the pattern consisting of concentric black and white circles. The purpose of the following experiments is to test this hypothesis.

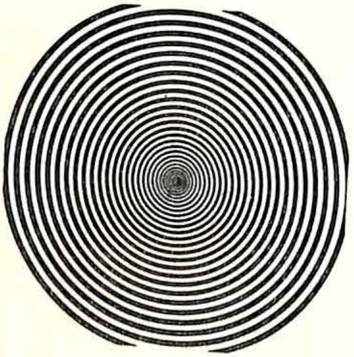
EXPERIMENT I

In this experiment the distance between the pattern and the observer is varied and the subject is requested to compare the amplitude of the motion of the radial fluctuating lines and their frequency at different distances. Campbell and Robson (1958) have already pointed out that the effect disappears when the eye is focused at infinity.

Subjects

Five subjects (S.M., K.F., M.H., M.M., C.S.) with an acuity of 20/20 participated in these observations. They were between 23 and 29 years of age and familiar with visual experiments.

FIGURE 1



Apparatus

The concentric ring illusion is shown in Figure 1. It was projected by a Carousel projector (Kodak Co.) onto a white screen. A camera shutter was attached to the objective. The room was kept dim. The luminance on the screen was 8 ft. L. as measured with an S.E.I. photometer. In order to keep the visual angle subtended by the pattern approximately constant, the subject and the projector were left next to each other and neither was moved throughout the experiment. The pattern subtended an angle of 7°

at the cornea. The width of a line was about 7 min. towards the periphery and decreased progressively to 2 min. of arc in the centre. The screen only was moved at the three distances used; namely, 4 m., 2 m. and 1 m. The luminances at these three distances were equalised using the diaphragm incorporated in the camera shutter.

The subjects wore their corrective lenses if any. The left eye was covered and an artificial pupil of 3 mm. was placed in front of the right eye. If the subject did not wear glasses, a trial frame was used.

Procedure

The screen was placed at any of the three distances and the subject shown the pattern. The subject was asked whether he noted the propeller effect of the pattern and to note if possible, the amount of motion (a quarter of a circle, a half of a circle, etc.) of the fluctuating radial lines and their frequency by pressing a counter each time a line fluctuated, per 10 sec. of observation. The screen was then placed at another distance, focused, the diameter of the diaphragm altered accordingly to keep the luminance constant and the procedure repeated. Two series of observations were thus carried out with each subject. They were requested to lean their heads against the wall behind and to remain as steady as possible during the observations.

Results

All five subjects noted that the effect was more marked, i.e. the motion and the frequency of the fluctuating radial lines were greatest at 1 m., less at 2 m., and much less at 4 m., but the effect was still present at 4 m. At 1 m. one subject for whom the effect appeared more regularly noted that the motion of the lines seemed to cover half a circle although sometimes the whole circle and the frequency was established to be of the order of 1 c/s. by having the subject count each time the line moved in the same direction. At 2 m., it appeared to be of the order of 0·5 c/s. and of the order of 0·25 c/s. at 4 m. These values are very approximate as it is very difficult to observe such a pattern steadily, and if fixation is maintained in the centre of the target after a few seconds Troxler's phenomenon (disappearance effect, Troxler, 1804) interferes with the observation. Thus most observations were made with free fixation to avoid this phenomenon occurring. However it was not possible to obtain quantitative results except to note the amplitude of motion of a line in terms of a half or a quarter of a circle and the observation that the fluctuating lines rotated more often. What made it impossible for the subjects to record the frequency pressing a key was the fact that a line may appear in one meridian then disappear and another appear in another meridian and this occurred quite irregularly in time and space.

With two of the five subjects a series of observations were carried out while biting a dental impression. The effect was compared with the situation in which the subject held his head steadily. The effect was less marked when biting a dental impression and it must be implied that slight head movements increase the effect. With the head maintained by the bite board the subject looked at the target placed at 33 cm., thus inducing 3 diopters of accommodation. Negative lenses (−1·00, −2·00, −3·00, −4·00) were placed randomly in front of one eye and the subject asked to compare the effect with the standard, that is without the negative lenses. The results showed that the effect was greater the more powerful the lens, or the larger the accommodation stimulus.

EXPERIMENT II

In this experiment, the accommodative system of the eye is paralysed, and its influence on the concentric ring illusion is thus assessed. Campbell and Robson (1958) have noted that the effect disappears when paralysing the eye with homatropine eye drops. In this experiment a more effective cycloplegic is used (Mitchell, 1959).

Subjects

Four of the subjects (S.M., K.F., M.M., C.S.) who took part in the first experiment were used.

Apparatus

The concentric ring illusion was projected by the same system as in Part I. The accommodation of the eye was paralysed by one drop in each eye of Cyclogyl 0·5 per cent. (Cyclopentolate hcl.). Accommodation was measured before instillation of the drug and again after 45 min. After that time the mean accommodation for the four subjects was 1·2 diopter. The accommodation was measured monocularly with an artifical pupil of 3 mm., using 20/20 letters moved towards the eye and of a luminance of 13·6 ft. L. Thus, the remaining amplitude of accommodation was sufficient to let the subject observe the pattern at 1 m.

Procedure

Part II of the study was carried out just after Part I so that a comparison could also be made between the observations made before and after paralysis of the accommodation. The same procedure as in Experiment I was employed. The experiment was carried out about 50 min. after instillation of the drug as it is known that its effect is optimum after that time (Mitchell, 1959).

Results

All four subjects noted a very marked reduction in amplitude of the motion and frequency of the concentric ring illusion at all distances. With three of the subjects who have very lightly pigmented eyes, there was no more propeller effect at 2 m. and at 4 m. and for one of them no effect at the three distances. This result can be accounted for by the fact that the drug acts most effectively on persons with lightly pigmented eyes (Mitchell, 1959).

EXPERIMENT III

The influence of accommodation can best be assessed when it is not present, as in aphakic subjects, particularly monocular aphakic.

Subject

The subject was one young male of 14 who had been operated for cataract one year ago in on eye due to an accident. He was wearing contact lenses in both eyes and had a visual acuity of 20/20 in each eye. The subject had no experience with visual experiments.

Apparatus

The same apparatus as in the other experiments was used. The distance was kept constant at 1 m. in front of his eyes since the depth of focus produced by the artifical pupil provided him with perfect vision at that distance with his aphakic eye. This experiment was carried out in the subject's home.

Procedure

The concentric ring illusion was first shown to the normal eye. The subject noted the effect. Then this eye was covered and the artifical pupil placed in front of the other eye. The procedure was repeated several times.

Results

The propeller effect was easily noted with his normal eye. He did not see the effect with his aphakic eye.

EXPERIMENT IV

In this experiment the exposure time of the concentric ring illusion is varied from short exposure of 1/100 sec. to a steady exposure of 5 sec.

Subjects

Five subjects (S.M., M.M., B.W., C.S., C.G.) took part in this experiment. They all have a visual acuity of 20/20 and are familiar with visual experiments.

Apparatus

The same apparatus as in the other experiments was used. The screen was placed 1 m. in front of the eyes. Short exposures were compensated by the use of higher luminance values in accordance with Block's law so as to achieve approximately equal intensities of stimulation. The background on which the concentric ring illusion was projected was brighter than in the other experiments; it had a luminance of 12 ft. L. and the pattern had a luminance of 100 ft. L. On the area of the screen where the centre of the pattern appeared, a small fixation point was drawn to maintain fixation before projecting the pattern.

Procedure

The subject was adapted for a few minutes before beginning the experiment. He was then shown the concentric ring illusion 5 times with a 10-sec. interval at each exposure time from 1/100 to 5 sec. The experiment was repeated in the reverse direction. In the third series of observations, the exposures were presented at random. The exposures used were 1/100, 1/25, ½, 1, 1·5, 2·5, 3 and 5 sec. The exposures beyond 1 sec. were obtained by using a stop watch calibrated in 1/10 of a sec. and pressing the trigger of the camera shutter on and off. The subject observed the pattern monocularly with an artifical pupil of 3 mm. and keeping his head as steady as possible during the observations by leaning against the wall or resting it on his arm.

Results

None of the subjects saw the propeller effect for exposures between 1/100 and ⅕ sec. They all noted it for 1 sec. but with a sensation that the illusion was only partial, at least in the amplitude of the motion. However, one subject did note the effect (partial) at ½ sec. but it appeared normal at 1 sec. For the other four subjects the concentric ring illusion appeared normal at 2 sec. for three of them and 1·5 sec. exposure for one of them.

Discussion and Conclusion

The results of these experiments demonstrate the role of the accommodative system in the concentric ring illusion. In Experiment I, it was shown that the motion and the frequency of the illusion increased as the pattern moved closer and when negative lenses were placed in front of the eye. Since contrast, visual angle, luminance and pupil size were approximately constant, it is unlikely that any other factors varied except the accommodation of the eye. Arnulf and Dupuy (1960) and Campbell *et al.* (1959) have demonstrated that the microfluctuations of accommodation increase as a function of accommodation and Arnulf and Dupuy (1960) proved that these microfluctuations were not equal in all the meridians of the eye but were astigmatic. These two factors account for the appearance of the increased sensation of rotating sectors within the pattern as its distance to the eye diminishes, and when accommodation is increased by negative lenses. The frequency of the involuntary eye movements such as the tremor is of the order of 30 to 70 c/s. (Ratliff and Riggs, 1950) which is well beyond that observed when viewing the concentric ring illusion. The frequency of occurrence of the illusion is approximately the same as that of the microfluctuations of accommodation, a fact already noted by Campbell and Robson (1958). The other involuntary eye movements such as the drifts and saccades are slower and of an order of magnitude comparable to that of the concentric ring illusion. But Mackay (1957*b*) who stabilized the retinal image of the concentric ring illusion thus eliminating the involuntary eye movements, and leaving the microfluctuations of accommodation, found that the illusion was unaffected. Involuntary eye movements do not make a major contribution to the phenomenon. The effect of eye movements would only shift the pattern in space rather than produce local variations within it.

According to Campbell *et al.* (1960) and Arnulf and Dupuy (1960), the microfluctuations of accommodation are eliminated by homatropine and this obviously extends to other cycloplegics such as cyclogyl which is a much more effective drug. However, it is not proven that they were completely eliminated since they might be below the noise level of their measuring instruments. This possibility is supported by the fact that some accommodation is still present with either drug and could explain why in some of the subjects of this experiment the concentric ring illusion is not completely absent after instillation of the drug. It must also be kept in mind that the instillation of a drug into the eye is open to error since some subjects blink more and lose some of the contents through the lacrymal canal; for example, in one subject, the course of pupil recovery was much longer in one eye than in the other.

The elimination of the microfluctuations of accommodation is more convincing when the eye has been operated for cataract. Although it was only possible to obtain one subject with monocular aphakia and the conditions of the experiment were not optimum, the result of this experiment makes it difficult indeed to account otherwise for the concentric ring illusion than by the microfluctuations of accommodation since both eye movements and central factors are unaffected.

In the fourth experiment, the exposure time was varied between 1/100 and 5 sec. The image is practically stabilized at 1/100 and no propeller effect is seen but at $\frac{1}{5}$ sec. most involuntary eye movements are present and yet the concentric ring illusion is still absent. This result is in agreement with Mackay's experiment (1957*b*) who found that the concentric ring illusion is unaffected when stabilized on the retina by means of an optical system. The speed at which the concentric ring illusion appears normal is between 1 and 2 sec. for most subjects and is more in accord with the speed of the microfluctuations of accommodation than with most eye movements.

Another series of observations were also carried out with the pattern placed at 1 m. A prism of 12 diopters, base down, was placed in front of one eye and the pattern looked at binocularly and therefore equally adapted. The subject could thus compare simultaneously the image of each eye separated in space, one above the other. Five subjects observed the concentric ring illusion in this fashion. Two of them had an anisometropia (a difference in the refractions of each eye) of more than 1 diopter. It is known that the accommodation in such cases may not be equal (Ball, 1952). When such is the case, it could be anticipated that the concentric ring illusion would appear somewhat different with each eye. This was confirmed for the two subjects who are anisometropic and in whom both eyes are healthy with no amblyopia. This result provides one more point in support of the role of the microfluctuations of accommodation as the underlying basis of the concentric ring illusion. Moreover, if central factors were responsible, this result could not have been obtained.

On the basis of the findings obtained in this study on the one hand and the evidence provided in the introduction by the experiments of Mackay (1957*b*) and Evans and Marsden (1966) on the other hand, it seems quite clear that the concentric ring illusion can be accounted for by the microfluctuations of accommodation and it is unnecessary to invoke either involuntary eye movements or central factors.

I thank Professor Lorrin A. Riggs for his encouragement throughout this study and for reading this manuscript, Edward Troendle, O.D. for having introduced me to one of his aphakic patients and the graduate students of the Laboratory who have so kindly served as subjects.

REFERENCES

ARNULF, A., DUPUY, O., and FLAMANT, F. (1951). Les microfluctuations d'accommodation de l'oeil et l'acuité visuelle pour les diamètres pupillaires naturels. *C.r. hebd. séanc. Acad. Sci. Paris*, **232**, 349.

ARNULF, A., and DUPUY, O. (1960). Contribution à l'étude des microfluctuations d'accommodation de l'oeil. *Revue Opt. theor. instrum.*, **39**, 195–208.

BALL, E. A. W. (1952). A study in consensual accommodation. *Amer. J. Optom.*, **29**, 561–74.

CAMPBELL, F. W., and ROBSON, J. G. (1958). Moving visual images produced by regular stationary pattern. *Nature*, **181**, 362.

CAMPBELL, F. W., and ROBSON, J. G. (1959). High-speed infrared optometer. *J. opt. Soc. Amer.*, **49**, 268–72.

CAMPBELL, F. W., WESTHEIMER, G., and ROBSON, J. G. (1959). Fluctuations of accommodation under steady viewing conditions. *J. Physiol. Lond.*, **145**, 579–94.

EVANS, C. R., and MARSDEN, R. P. (1966). A study of the effect of perfect retinal stabilization on some well-known visual illusions using the after-image as a method of compensating for eye movements. *Brit. J. Physiol. Opt.*, **23**, 242–8.

HELMHOLTZ, H. (1867). *Handbuch der Physiologischen Optik*. Leipzig: Leopold Voss.

LE GRAND, Y. (1965). *Optique physiologique*, Vol. 1, 3rd edition. Paris: Revue d'Optique.

MACKAY, D. M. (1957*a*). Moving visual images produced by regular stationery patterns. *Nature*, **180**, 849–50.

MACKAY, D. M. (1957*b*). Some further visual phenomena associated with regular patterned stimulation. *Nature*, **180**, 1145–6.

MITCHELL, D. W. A. (1959). *The Use of Drugs in Refraction*. London: The British Optical Association.

PRITCHARD, R. M. (1958). Visual illusions viewed as stablized retinal images. *Quart. J. exp. Psychol.*, **10**, 77–81.

RATLIFF, F., and RIGGS, L. A. (1950). Involuntary motions of the eye during monocular fixation. *J. exp. Psychol.*, **40**, 687–701.

TROXLER, D. (1804). Über das Verschwinden gegebener Gegenstände innerhalb Unseres GesichtsKreises. *Opthalmologische Bibliothek*. Vol. 2. Jena: Himly und Schmidt.

Manuscript received 3rd May, 1968.

PERCEPTUAL INDEPENDENCE IN THE PERCEPTION OF SPEECH

BY

D. W. J. CORCORAN, D. D. DORFMAN and D. L. WEENING

From Medical Research Council, Applied Psychology Research Unit, Cambridge, and San Diego State College

The purpose of the experiments and analyses was to discover whether the "independence" model, which successfully accounts for the perception of synthetic multi-dimensional stimuli, would also describe the perception of speech in noise. Two sets of four words were chosen which could be considered phonetically to be describable in terms of two dimensions. The S/R matrix was analysed as if the stimuli were bi-dimensional elementary stimuli. Satisfactory fits were obtained by the independence model. Two analyses were conducted upon data from Miller and Nicely (1955) and Conrad (1964) to discover whether the perception of phonemes was also predictable on the independence model. In general this was found to be the case.

INTRODUCTION

The perception of speech exhibits features which suggest that its underlying perceptual mechanism may be in some way different in principle from that which processes other auditory inputs. It is probably the largest auditory input, and the major initiator of changes in behaviour. It has "meaning." It can be distorted by peak clipping, filtering, masking, reverberating, by phase distortion, time sampling or by passage through non-linear circuits and still remain intelligible to a degree which is not perfectly related to the purely physical changes which have been induced by the methods (Licklider and Miller, 1951). It would appear to be processed mainly by the left hemisphere, whilst the major portion of other auditory inputs are apparently shunted through the right (e.g. Milner, 1962; Kimura, 1961*a*, *b*, 1964; Chaney and Webster, 1965). The storage of its "items" is immense (Oldfield, 1966) and, perhaps most significant of all, the perceiver can also *generate* stimuli—a fact which some theorists (e.g. Liberman, 1957; Liberman *et al.*, 1963; Stevens, 1960) consider of paramount importance in understanding its perception. It has therefore become fashionable to consider the mechanism of speech perception to have evolved along unique principles.

One way to decide whether speech perception is fundamentally similar or dissimilar to that of other auditory inputs is to discover whether a successful model for one type of input is equally meaningful for the other. This is the purpose of the following analyses and experiments.

The non-speech system

Perhaps the simplest and most successful models to describe the perception of multi-dimensional elementary stimuli have been based on the hypothesis that the physically defined parameters are dealt with independently by the perceptual system (e.g. Eriksen and Hake, 1955; Corcoran, 1966, 1967). The separate experiments require somewhat different applications of the model, and a generalized description is best framed in terms of the 1967 experiment, whose paradigm is identical to those to be reported here.

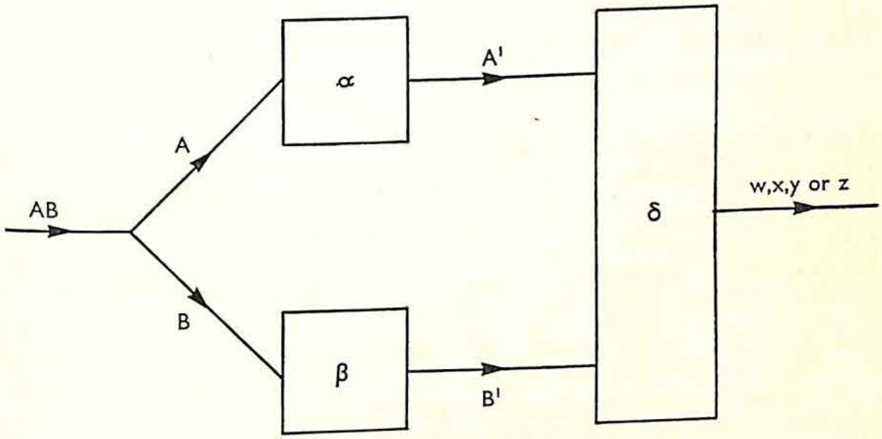
In the 1967 experiment four auditory stimuli were synthesised from two frequencies (A_1 and A_2) and two degrees of amplitude modulation (B_1 and B_2). Stimulus A_1B_1 was a 1,000 Hz pure tone modulated three times per second at 60 per cent. A_1B_2 was also 1,000 Hz, but was unmodulated; A_2B_1 was a 1,250 Hz modulated tone and A_2B_2 a 1,250 Hz unmodulated tone.

After training which ensured that they could be reliably named, W, X, Y and Z, the stimuli were presented in noise for identification.

It was reasoned that if an observer's judgment about one parameter were uninfluenced by the particular value taken by the other (i.e. that he was, for example, as likely to be correct in judging A_1 to be A_1 given A_1B_1 as he was when given A_1B_2) then the probability of perceiving the whole stimulus correctly would be simply the product of the probabilities of perceiving its "parts" correctly. Thus the probability of correctly perceiving A_1B_1 for example would be pa_1pb_1 where pa_1 is the probability that A_1 will be perceived as A_1 and pb_1 the probability that B_1 will be perceived as B_1.

Within this hypothesis is buried two assumptions (1) that A and B are *sensorily* independent, and (2) that the perceptual system combines the discreet information it receives from the analysers multiplicatively. A system which would behave in this fashion is illustrated in Figure 1, which shows two "analysers" α and β which receive A and B respectively and whose function is to resolve A_1 from A_2 and B_1 from B_2. The outputs A'_1, A'_2, B'_1 and B'_2 feed into a decision maker δ which initiates output W, if it receives $A'_1B'_1$, X if it receives $A'_1B'_2$, Y if it receives $A'_2B'_1$ and Z when given $A'_2B'_2$.

FIGURE 1



Illustrative diagram of a system which would operate on the principle of perceptual independence.

If pa_1 is the probability that α will give output A'_1 given A_1, pa_2 the probability that α will yield A'_2 given A_2, pb_1 the probability that β will yield B'_1 given B_1 and pb_2 the probability that β will give output B'_2 given B_2, then the probability that δ will signal W given stimulus A_1B_1 will be simply pa_1pb_1. All 16 S–R contingencies are predictable and each is shown in Table I. Cell 3, 2 for example predicts that the probability that stimulus A_2B_1 will be named X is the product of the probabilities that α will incorrectly give output A'_1 given A_2, and β will yield B'_2 given B_1. Thus δ will receive $A'_1B'_2$ when in fact A_2B_1 had been presented. It is notable that all errors in the system can be traced to failures of α and β to discriminate their inputs.

Each cell of Table I may contain experimentally determined probabilities, merely by presenting the four stimuli (preferably in noise) and requiring subjects to name each stimulus W, X, Y or Z. Thus 16 equations can be formed with which to estimate the four unknown underlying probabilities pa_1, pa_2, pb_1 and pb_2. These estimates were found to agree well in the 1967 experiment and it was concluded that the model had provided a good fit to the data.

The system described above does not, however, accord well with contemporary theoretical preferences, which are more in favour of an underlying continuous rather than a discrete process (Swets, 1964; Green and Swets, 1966). Accordingly, one might postulate that α and β function as "signal detectors." Let us therefore assume A_1 and A_2 to be two unidimensional stimuli located on dimension A. For simplicity assume that A_2 is normally distributed with a mean of zero and unit variance and that A_1 is also normal with unit

TABLE I

STIMULUS-RESPONSE MATRIX INDICATING THE CONTINGENT PROBABILITIES UNDERLYING RESPONSES W, X, Y AND Z TO A_1B_1, A_1B_2, A_2B_1 AND A_2B_2

	Stimulus			
Response	1 A_1B_1	2 A_1B_2	3 A_2B_1	4 A_2B_2
1 W (A_1B_1)	pa_1pb_1	$pa_1(1-pb_2)$	$(1-pa_2)pb_1$	$(1-pa_2)(1-pb_2)$
2 X (A_1B_2)	$pa_1(1-pb_1)$	pa_1pb_2	$(1-pa_2)(1-pb_1)$	$(1-pa_2)pb_2$
3 Y (A_2B_1)	$(1-pa_1)pb_1$	$(1-pa_1)(1-pb_2)$	pa_2pb_1	$pa_2(1-pb_2)$
4 Z (A_2B_2)	$(1-pa_1)(1-pb_1)$	$(1-pa_1)pb_2$	$pa_2(1-pb_1)$	pa_2pb_2

variance and has a mean of d'_A. Assume also that B_1 and B_2 are similarly located on dimension B such that B_2 has a zero mean and B_1 a mean of d'_B and that B_1 and B_2 are normally distributed with unit variance. Finally, assume that X_A and X_B are continuous random variables, where X_A is an observation on dimension A and X_B an observation on B and that when a compound two dimensional stimulus is received, an observation of the form (X_A, X_B) results. Four decision rules are necessary to convert the sensory events into a discrete response. These are: (1) If $X_A > K_A$ and $X_B > K_B$ decide A_1B_1. (2) If $X_A > K_A$ and $X_B < K_B$, decide A_1B_2. (3) If $X_A < K_A$ and $X_B > K_B$, decide A_2B_1 and (4) If $X_A < K_A$ and $X_B < K_B$, decide A_2B_2. K_A and K_B are cut-off points on dimensions A and B respectively (see Fig. 2). It can be easily proved that the theoretical probabilities of each S–R contingency are those presented in Table II, where F is the cumulative normal function. The theory states in effect that the quadrant of Figure 2 in which the compound observation falls will determine the response. It is interesting to note that if we let $pa_1 = F(d'_A - K_A)$, $pa_2 = F(K_A)$, $pb_1 = F(d'_B - K_B)$ and $pb_2 = F(K_B)$, one is left with the simple forementioned independence model.

TABLE II

STIMULUS-RESPONSE MATRIX INDICATING THE PREDICTIONS FROM THE CONTINUOUS MODEL

	Stimulus			
Response	1 A_1B_1	2 A_1B_2	3 A_2B_1	4 A_2B_2
1 W (A_1B_1)	$F(d'_A-K_A)F(d'_B-K_B)$	$F(d'_A-K_A)F(-K_B)$	$F(-K_A)F(d'_B-K_B)$	$F(-K_A)F(-K_B)$
2 X (A_1B_2)	$F(d'_A-K_A)F(K_B-d'_B)$	$F(d'_A-K_A)F(K_B)$	$F(-K_A)F(K_B-d'_B)$	$F(-K_A)F(K_B)$
3 Y (A_2B_1)	$F(K_A-d'_A)F(d'_B-K_B)$	$F(K_A-d'_A)F(-K_B)$	$F(K_A)F(d'_B-K_B)$	$F(K_A)F(-K_B)$
4 Z (A_2B_2)	$F(K_A-d'_A)F(K_B-d'_B)$	$F(K_A-d'_A)F(K_B)$	$F(K_A)F(K_B-d'_B)$	$F(K_A)F(K_B)$

FIGURE 2

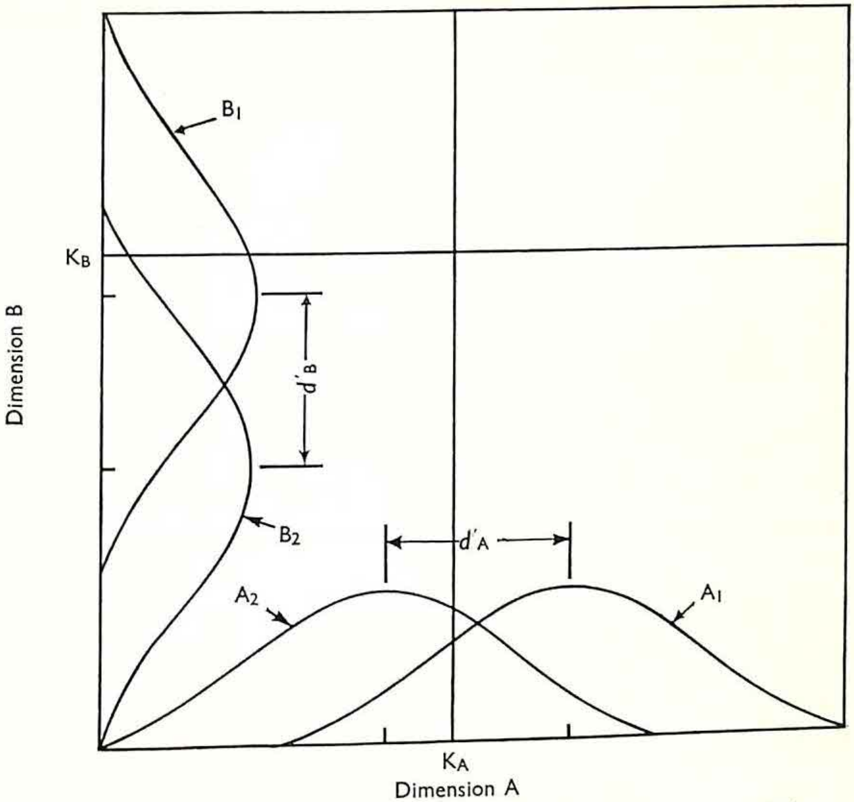


Illustration of the continuous version of the independence model.

For purposes of fitting the model to the data, maximum likelihood estimates of the parameters d'A, KA, d'B, KB, may be obtained. It can be shown that:

$$\hat{p}a_1^{-1} = \hat{F}^{-1}(d'_A - K_A) = \hat{d}'_A - \hat{K}_A = \left\{ \frac{\sum\limits_{A_1 \wedge A'_1} Nij}{\sum\limits_{A_1 \wedge A'_1} Nij + \sum\limits_{A_1 \wedge A'_2} Nij} \right\}^{-1} \quad .. \quad (1)$$

$$\hat{p}a_2^{-1} = \hat{F}^{-1}(K_A) = \hat{K}_A = \left\{ \frac{\sum\limits_{A_2 \wedge A'_2} Nij}{\sum\limits_{A_2 \wedge A'_1} Nij + \sum\limits_{A_2 \wedge A'_2} Nij} \right\}^{-1} \quad .. \quad (2)$$

$$\hat{p}b_1^{-1} = \hat{F}^{-1}(d'_B - \hat{K}_B) = \hat{d}'_B - \hat{K}_B = \left\{ \frac{\sum\limits_{B_1 \wedge B'_1} Nij}{\sum\limits_{B_1 \wedge B'_1} Nij + \sum\limits_{B_1 \wedge B'_2} Nij} \right\}^{-1} \quad .. \quad (3)$$

and $$\hat{p}b_2^{-1} = \hat{F}^{-1}(K_B) = \hat{K}_B = \left\{ \frac{\sum\limits_{B_2 \wedge B'_2} Nij}{\sum\limits_{B_2 \wedge B'_1} Nij + \sum\limits_{B_2 \wedge B'_2} Nij} \right\}^{-1} \quad .. \quad (4)$$

where A_K and B_K are stimulus components, A'_K and B'_K are response components, Nij is the number of observations in the i^{th} row and j^{th} column of the stimulus-response matrix, F^{-1} is the inverse of the normal distribution function, and the summation is over all events in the matrix containing the event under the summation. Note (1), (2), (3) and (4) are four equations in four unknowns, and the solution of these equations gives maximum likelihood estimates of d'_A, K_A, d'_B, and K_B, and therefore maximum likelihood estimates of pa_1, pa_2, pb_1, and pb_2. However, these estimates *are identical to those obtained with the simple independence model, and the predictions of the equivalent values in a matrix such as Tables I or II are therefore exactly the same for the two models*. Thus whichever of the two models is preferred for aesthetic or other reasons, the predictions are the same, and are, to judge from the earlier experiments, good predictions of the confusions amongst synthetic stimuli.

The speech system

Although words are describable in terms of measurable patterns in time, amplitude and frequency, the problem of how the separate acoustic dimensions are combined into reliable percepts is very complex (e.g. Liberman, 1957), and no discussion will be attempted here. Instead we shall consider words to be built from phonemes, and select groups of words where phonetic characteristics differ in terms of a limited and specifiable number of (phonetic) dimensions. This cannot be done with any great accuracy, but between words such as SAD and MAD, there is probably little difference other than the first phoneme.

The present experiments have therefore *defined* a word such as SAD as *logically equivalent* to the synthetic stimulus A_1B_1. Instead of a difference in pure frequency between A_1 and A_2, the difference between SAD and MAD is the difference between that *pattern* of physical values that distinguishes an /s/ from an /m/. We shall further define SAT as A_1B_2, and MAT as A_2B_2. When these words are presented in suitable levels of noise it is possible to obtain a matrix like those of Table I or II the cells of which will contain the probabilities that each presented word will be named each of the possible words in the specified vocabulary of four. We many then ask: are these probabilities predictable on the same model as the equivalent synthetic stimuli?

EXPERIMENT I

Method

Stimulus material. The words SAD, SAT, MAD and MAT were recorded in random order with the restriction that each occurred with an equal overall frequency within the 128 item recording. The words were spoken by a "standard" American voice with 5 sec. intervals between words. The second channel of the magnetic tape contained filtered thermal noise, whose 3 dB. down points lay at 500 and 6,500 Hz. It was necessary to use filtered noise in order to make the resolution of /s/ from /m/ and /t/ from /d/ about equal. This was of no conceptual importance, merely a convenience. The speech and noise mixture was relayed over earphones at speech-to-noise (SN) ratios which varied somewhat between subjects. Since large individual differences were evident in the ability to listen to speech in noise, it was necessary to perform a crude psychophysical procedure before the test proper in order to set the signal and noise attenuators at what were considered to be suitable levels for each subject. In each instance the experimenter decided upon a value, which he considered would lead to about 60–70 per cent. correct recognition. This level of performance was considered likely to yield a convenient number of errors. Thus the rms. SN ratios varied between zero and −26 dB. with a median value of −20 dB.

Subjects and procedure. Seven undergraduates of San Diego State College were employed as subjects. They were instructed after adjustment of SN ratios, that they would hear only SAD, SAT, MAD and MAT, that these words would often be difficult to hear, but that they should nevertheless record an answer to each item.

Results

The upper values in the cells of Table III are the empirical probabilities that each of the four stimuli will be assigned to the four responses shown in the first column. Each probability is a mean value for each S/R contingency. The probabilities in brackets are the predictions from the models. The scores were treated individually and the recorded predictions are the mean of the individual predictions.

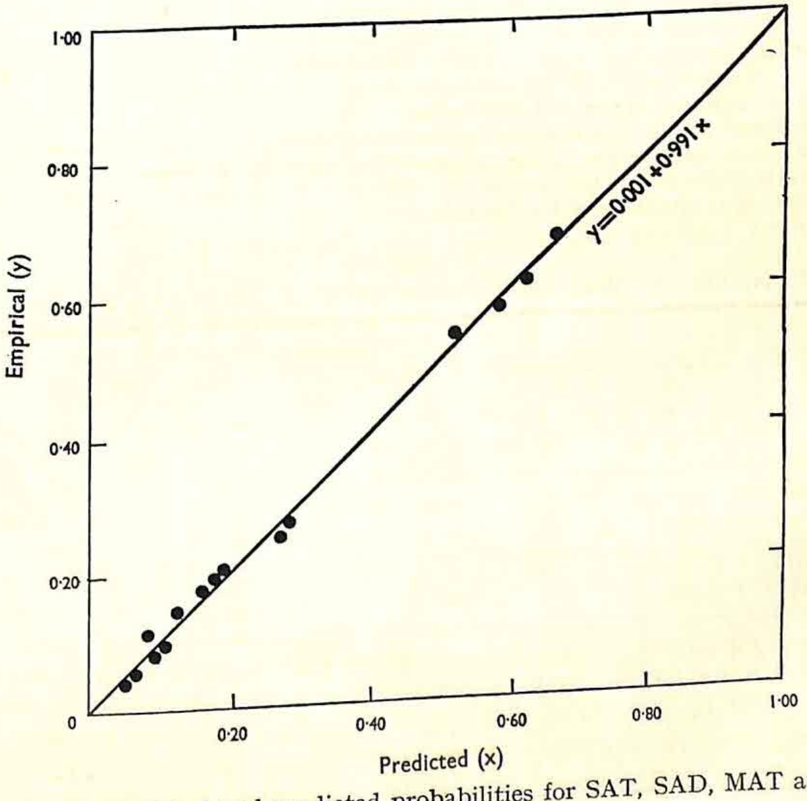
TABLE III

EMPIRICAL AND PREDICTED SCORES FOR SAD, SAT, MAD, MAT

	Stimulus			
Response	1 SAD A_1B_1	2 SAT A_1B_2	3 MAD A_2B_1	4 MAT A_2B_2
1 SAD (A_1B_1)	0·61 (0·62)	0·25 (0·27)	0·08 (0·09)	0·11 (0·08)
2 SAT (A_1B_2)	0·19 (0·18)	0·53 (0·52)	0·04 (0·05)	0·05 (0·06)
3 MAD (A_2B_1)	0·16 (0·16)	0·09 (0·10)	0·68 (0·67)	0·27 (0·28)
4 MAT (A_2B_2)	0·05 (0·06)	0·14 (0·12)	0·20 (0·19)	0·57 (0·58)

Maximum likelihood estimates of pa_1, pa_2, pb_1, and pb_2 were obtained from the empirical matrix of each subject, and these estimates were used to generate the predicted matrix for each subject by use of the equations in Tables I and II. Note that the model uses four parameters to predict 16 proportions, 12 of which are

FIGURE 3



Relation between empirical and predicted probabilities for SAT, SAD, MAT and MAD.

independent. The method of maximum likelihood requires all the data for estimation of parameters. In the 1967 experiment, the parameter estimates used to predict the data of a given cell, were obtained from the data of other cells. In practice, the difference in procedure makes little difference to the eventual prediction.

Figure 3 plots the empirical against the predicted scores and the line, which represents the best linear fit also functions on the line indicating perfect correspondence between the two variables. The mean error is +0·0006, indicating a negligible and quite insignificant over-prediction. The correlation between the two sets of scores is 0·991 and the regression equation, which should ideally be $Y = X$, is $Y = 0{\cdot}001 + 0{\cdot}991X$.

EXPERIMENT 2

Experiment 1 used four words which were necessarily similar phonetically, but which were also (perhaps unnecessarily) rather similar in their probability of occurrence in language. The words were in fact either A or AA on the Thorndike and Lorge count (1944), with the exception of MAT, which has a G rating of 18. The following experiment therefore uses words which differ more markedly in probability of occurrence, since it is well established (e.g. Broadbent, 1967) that the more probable a word the more likely it is to be perceived correctly in noise. It may, of course, be argued that word probability will have little or no effect when the population is small and known, but this still needs to be questioned in passing.

Method

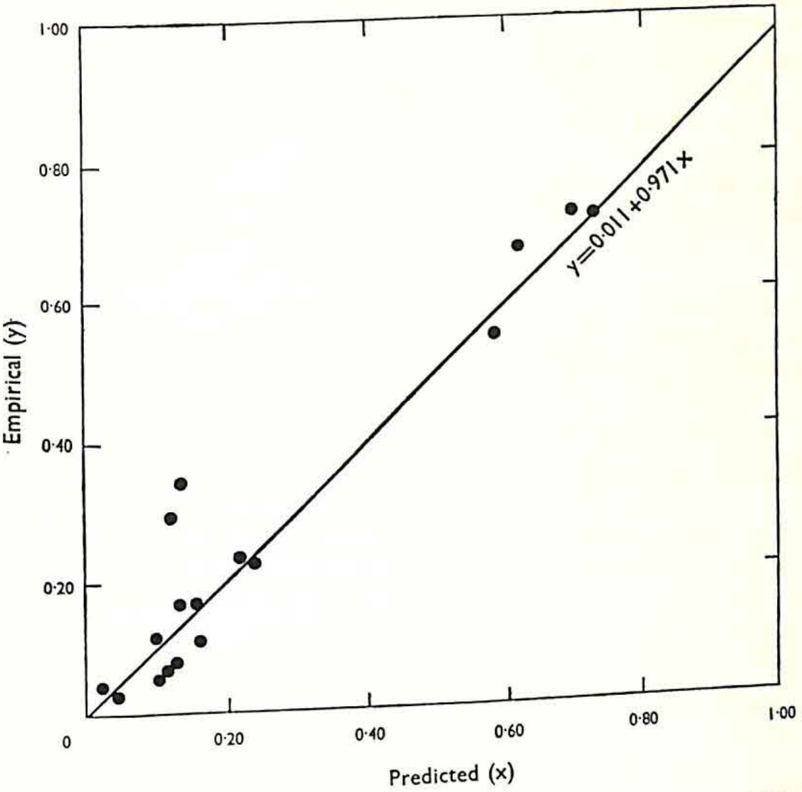
The words MUF, MUD, CUF and CUD were used as stimuli; these have G ratings of (MUFF) 5, (MUD) 47, (CUFF) 13 and (CUD) 1. They were recorded at 5 sec. intervals on one channel of a double channel tape and filtered noise with 3 dB. down points at 1,600 and 6,500 Hz was recorded on channel 2. The order of presentation was random with the restriction that each word occurred with equal frequency within the 128 item recording. The initial psychophysical procedures results in SN ratios from +2 dB. to −27 dB. with a median value of −18 dB. The method was otherwise identical to that of the previous experiment. Seven college students served as subjects and operated under instructions identical to those of Experiment 1.

The empirical scores and the predictions are presented in Table IV and the plot of empirical against predicted estimate is shown in Figure 4. The mean error is +0·0044; the correlation coefficient is 0·968, and the regression in equation $Y = 0{\cdot}011 + 0{\cdot}971X$ is again a close approximation to the ideal fit.

TABLE IV
EMPIRICAL AND PREDICTED MATRIX FOR MUF, MUD, CUF AND CUD

	Stimulus			
Response	1 MUF A_1B_1	2 MUD A_1B_2	3 CUF A_2B_1	4 CUD A_2B_2
MUF 1 (A_1B_1)	0·71 (0·73)	0·11 (0·16)	0·23 (0·22)	0·06 (0·05)
MUD 2 (A_1B_2)	0·17 (0·15)	0·72 (0·70)	0·03 (0·04)	0·23 (0·23)
CUF 3 (A_2B_1)	0·08 (0·12)	0·06 (0·05)	0·67 (0·62)	0·17 (0·13)
CUD 4 (A_2B_2)	0·04 (0·02)	0·11 (0·10)	0·07 (0·11)	0·54 (0·58)

FIGURE 4



Relation between empirical and predicted probabilities for MUF, MUD, CUF and CUD.

Discussion of Experiments 1 and 2

The two experiments have shown that the model predicts as well for words as for synthetic stimuli (cf. Corcoran, 1967). It therefore becomes of interest to re-examine the logic of using words as if they were a compound of a discreet number of independent phonetic dimensions.

In defining SAD as A_1B_1, SAT as A_1B_2, etc., two assumptions were made: (1) that *only* two dimensions discriminate the four words: dimension /s/–/m/ and dimension /d/–/t/. (2) that these dimensions were orthogonal, i.e. that the discrimination of /s/ from /m/ occurred independently of the discrimination of the later phoneme. Assumption (1) really states that only two *relevant* or *usable* dimensions were involved and that other factors like the overall stimulus amplitude, the stress upon certain phonemes, and the quality of the vowel, did not differ (or were randomly varied) amongst the words. Assumption (2) has more interesting implications, since it states in effect that the perception of a whole word is built up from *independent* decisions about the nature of its component phonemes. We may therefore think of word perception as built from functional components not unlike that shown in Figure 1, with the important difference that α and β must be considered to be programmed to discriminate between physical *patterns* (rather than single dimensional values) indicative of phonemes.

Analyses of Data by Miller and Nicely (1955) and Conrad (1964)

In the previous experiments there was some evidence to show that phonemes were the basic "units" from which the perception of the whole word was built. One may legitimately ask therefore, whether the perception of phonemes is built up in much the same way or whether the independence, multiplicative model does not apply.

Phonetically, spoken utterances have been classified in terms of the *anatomical pattern of movements* which the vocal apparatus assumes in the *production* of speech. (Miller and Nicely, 1955; Peterson and Shoup 1966). Miller and Nicely's description was confined to some sixteen consonants for which they defined five dimensions: Voicing, Nasality, Affrication, Duration and Place. All dimensions but Place were considered to be binary, the latter being subdivided into three categories: front, middle and back. (The Peterson and Shoup analysis was more discriminatory in defining 13 places, but here we shall use Miller and Nicely's classification, since we shall use their experimental data for analysis). Table V is a reproduction of Miller and Nicely's Table XIX, in which each of the consonants used in their study is described in terms of the anatomical dimensions. From the latter Table we selected for analysis two sets of four phonemes which correspond (at least logically) to the synthetic stimuli used in the 1967 study.* One set was /n/, /m/, /d/ and /b/ and the other /t/, /d/, /p/ and /b/; it will be noted that /d/ and /b/ are common to both sets. We shall arbitrarily define /n/ as logically equivalent A_1B_1, since it is Nasal (A_1) and its place of articulation is Middle (B_1); we shall consider /m/ to represent A_1B_2, since it is Nasal (A_1), but frontally articulated (B_2); /d/, on the other hand, is non-Nasal (A_2) but is middle (B_1); and is therefore A_2B_1, whilst /b/ is both non-Nasal and Frontal and will therefore be considered logically equivalent to A_2B_2. Similarly /p/ has been

TABLE V

A Reproduction of Miller and Nicely's Table XIX Classification of Consonants in Terms of Anatomical Characteristics of Speech Production

Consonant	*Voicing*	*Nasality*	*Affrication*	*Duration*	*Place*
p	0	0	0	0	0
t	0	0	0	0	1
k	0	0	0	0	2
f	0	0	1	0	0
o	0	0	1	0	1
s	0	0	1	1	1
ʃ	0	0	1	1	2
b	1	0	0	0	0
d	1	0	0	0	1
g	1	0	0	0	2
v	1	0	1	0	0
ð	1	0	1	0	1
z	1	0	1	1	1
	1	0	1	1	2
m	1	1	0	0	0
n	1	1	0	0	1

* Reproduced from *J. acoust. Soc. Amer.*, **27**, p. 347 by permission of the Editor and authors.

---

* It is, of course, possible to use the entire 16 × 16 matrices presented by Miller and Nicely, but this involves a time-consuming mathematical treatment.

defined as A_1B_1, since it is unvoiced and frontally articulated, /b/ as A_1B_2 since it is voiced and frontal, /t/ A_2B_1 since it is unvoiced and medially articulated and /d/ as A_2B_2 since it is both voiced and medially articulated. A similar classification can be applied to several other groupings of consonants. We have selected these two sets because they sample the five dimensions fairly well and because it is useful to replicate the behaviour of the same phonemes /d/ and /b/ within a different set. In equating these spoken stimuli to the electronically produced stimuli of the earlier experiments, we are assuming that the dimensions of Miller and Nicely's classification are orthogonal (sensorily independent), thus it will be assumed for example that although /m/ and /n/ differ in place of articulation, one will not be more (or less) *nasal* in quality than the other.

The necessary experiment to test the fit of the model has been conducted and reported in the Miller and Nicely study. Subjects were required to identify 16 consonants when spoken in noise: each consonant was spoken as the initial phoneme and was followed by /a/ as in father. The experiments were conducted under many varied conditions of masking, but we have selected probabilities from the matrix in which the SN ratio was −12 dB. and the filter set at 200 to 6,500 Hz, because this matrix has a liberal distribution of errors but is clearly non-random.

The probabilities of confusion between /n/, /m/, /d/ and /b/ and between /p/, /b/, /t/ and /d/ were extracted from Miller and Nicely's 16 × 16 matrix and normalized to yield the matrices shown in Tables VI and VII. In extracting 16 cells from the 156 cell matrix we have assumed that the probabilities of confusion between the stimuli are independent of the other stimuli in the experiment. This assumption may or may not be valid. The probabilities are therefore estimates of those which would have occurred if only /n/, /m/, /d/ or /b/ or /p/, /b/, /t/ and /d/ had been used as stimuli.

TABLE VI

DATA FROM MILLER AND NICELY /n/, /m/, /d/ and /b/

	Stimulus			
Response	1 /n/ A_1B_1	2 /m/ A_1B_2	3 /d/ A_2B_1	4 /b/ A_2B_2
1 /n/ (A_1B_1)	0·62 (0·62)	0·33 (0·25)	0·14 (0·17)	0·09 (0·07)
2 /m/ (A_1B_2)	0·35 (0·32)	0·60 (0·69)	0·10 (0·09)	0·19 (0·19)
3 /d/ (A_2B_1)	0·02 (0·02)	0·02 (0·04)	0·55 (0·49)	0·17 (0·20)
4 /b/ (A_2B_2)	0·02 (0·02)	0·06 (0·04)	0·21 (0·25)	0·56 (0·54)

Tables VI and VII may be equated with Tables I and II if the perception of phonemes is describable in terms of one of the independence models. Thus pa_1 becomes the probability that a nasal utterance will be considered to be nasal; pa_2 becomes the probability that a non-nasal utterance will be considered to be non-nasal pb_1 becomes the probability that the middle articulatory position will be correctly

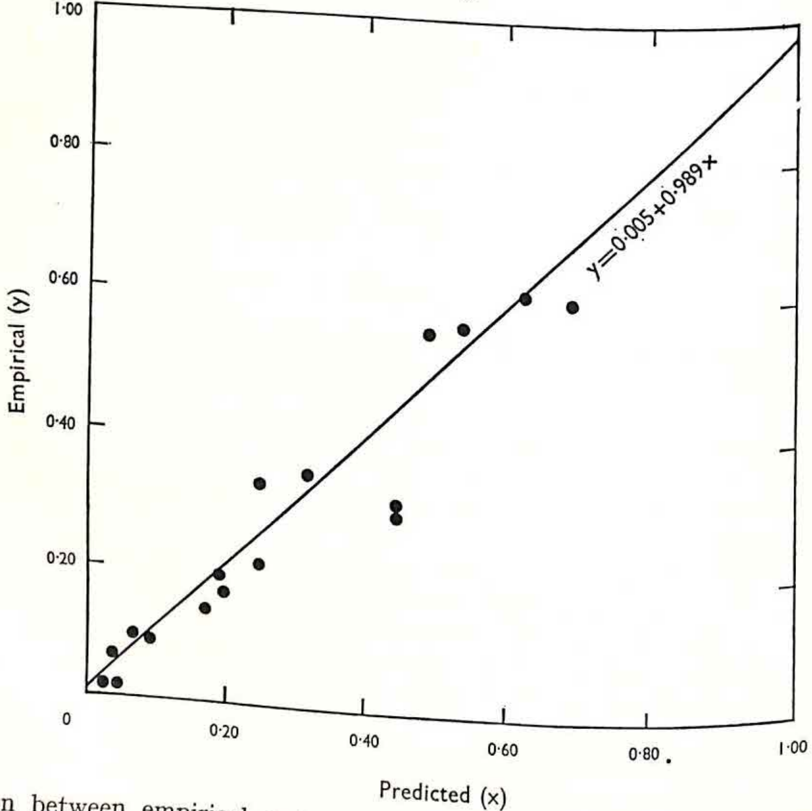
TABLE VII

DATA FROM MILLER AND NICELY /p/, /t/, /b/, /d/

	Stimulus			
Response	1 /p/ A_1B_1	2 /t/ A_1B_2	3 /b/ A_2B_1	4 /d/ A_2B_2
1 /p/ (A_1B_1)	0·42 (0·59)	0·47 (0·36)	0·02 (0·02)	0·00 (0·01)
2 /t/ (A_1B_2)	0·52 (0·37)	0·51 (0·60)	0·02 (0·02)	0·04 (0·03)
3 /b/ (A_2B_1)	0·03 (0·02)	0·02 (0·02)	0·73 (0·59)	0·26 (0·36)
4 /d/ (A_2B_2)	0·02 (0·01)	0·00 (0·03)	0·22 (0·37)	0·70 (0·60)

recognized and pb_2 becomes the probability that a non-middle (frontal) articulation will be perceived as such. The results for /n/, /m/, /d/ and /b/ are presented in Table VI and Figure 5. On this occasion, since the scores of individual subjects

FIGURE 5

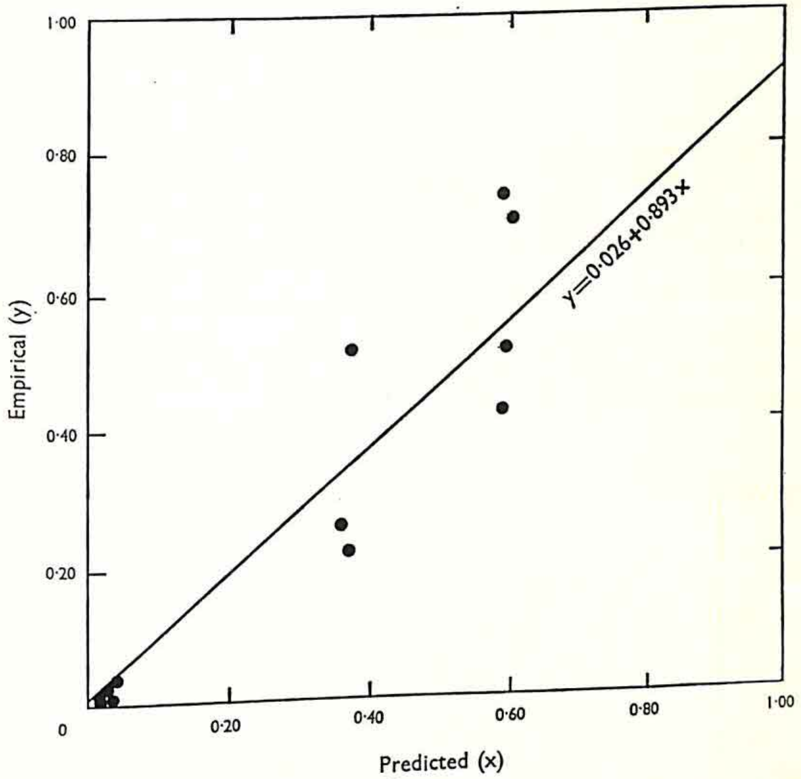


Relation between empirical and predicted probabilities for /n/, /m/, /d/ and /b/ (Miller and Nicely data).

were not available the computations were conducted upon the group means. Even so the empirical scores are satisfactorily predicted. The mean error was −0·0019, indicating a small underprediction, the correlation between the readings was 0·989 and the regression equation read Y = 0·005 + 989X.

The empirical values for /p/, /b/, /t/ and /d/, however, are poorly predicted (Table VII, Figure 6). But the empirical scores show a feature which suggests that the original matrix may be unreliable; Phoneme /t/ is more frequently heard as /p/ than as /t/ and vice versa (see cells 1, 1; 2, 1; 1, 2; 2, 2). The mean error was still small, −0·0013, but the correlation was lower 0·868, and the regression equation was a poor approximation of Y = X, i.e. 0·026 + 0·893X.

FIGURE 6



Relation between empirical and predicted confusions amongst /p/, /t/, /b/ and /d/ (Miller and Nicely data).

There are several sources from which this "negative" finding may have sprung (*a*) Miller and Nicely's classification may be untrue, (*b*) voicing and place may not be orthogonal, e.g. that the voicing associated with the production of a /d/ is different from that associated with that of a /b/, or (*c*) that the models are untrue.

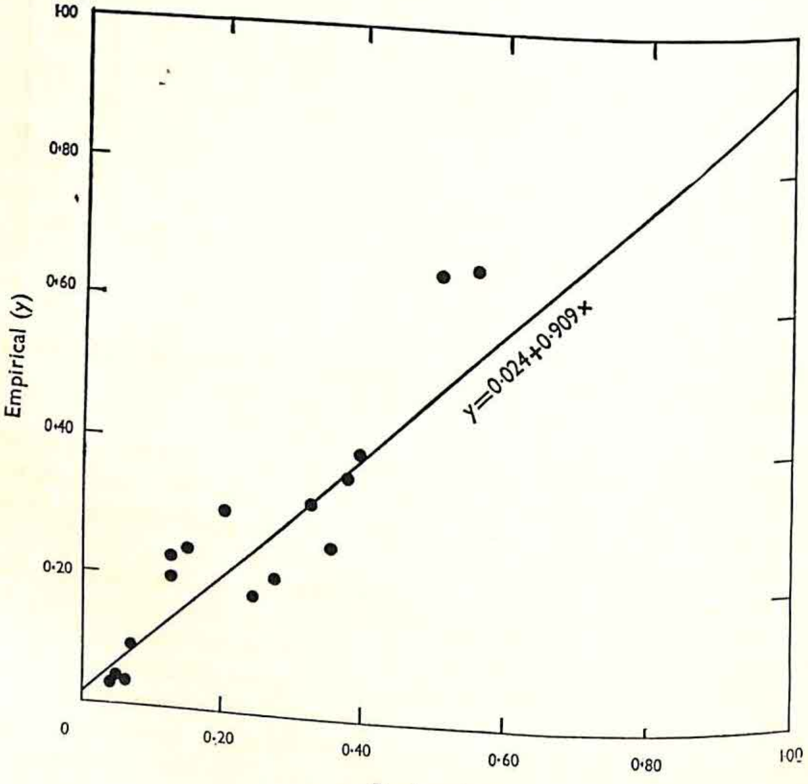
Table VIII however, suggests that the cause of the delinquent behaviour of the Voicing dimension was likely to be some factor in Miller and Nicely's *experiment*. This table contains estimates of the same underlying probabilities taken from data by Conrad (1964). Conrad presented letter *names* in noise and recorded the confusion matrix between the 26 letters. The stimuli "tee," "dee," "pee" and "bee" are similar to Miller and Nicely's "ta," "da," "pa" and "ba" and the same analysis is

TABLE VIII

DATA FROM CONRAD /p/, /t/, /b/, /d/

	Stimulus			
Response	1 /p/ A_1B_1	2 /t/ A_1B_2	3 /b/ A_2B_1	4 /d/ A_2B_2
1 /p/ (A_1B_1)	0·56 (0·56)	0·55 (0·51)	0·23 (0·23)	0·30 (0·21)
2 /t/ (A_1B_2)	0·32 (0·33)	0·36 (0·38)	0·20 (0·13)	0·24 (0·15)
3 /b/ (A_2B_1)	0·09 (0·07)	0·04 (0·06)	0·39 (0·34)	0·26 (0·36)
4 /d/ (A_2B_2)	0·04 (0·04)	0·05 (0·05)	0·18 (0·24)	0·21 (0·28)

FIGURE 7



Relation between empirical and predicted confusions amongst /p/, /t/, /b/ and /d/ (Conrad's data).

possible. The overall probabilities differ, of course, due to differences in speech to noise ratio, between the experiments.

The analysis of Conrad's data is shown in Table VIII and Figure 7. Here there is obviously a high correspondence between empirical observation and prediction. Mean error $-0{\cdot}0013$, $r = 0{\cdot}912$, and the regression equation was of the form $Y = 0{\cdot}024 + 0{\cdot}909X$.

The analyses of Miller and Nicely's and Conrad's data therefore suggest that on the whole the anatomical dimensions are good descriptions of the underlying perceptual "units," that they are probably independent and that as a result the confusions amongst phonemes are predictable on the multiplicative model.

General Discussion

Referring to the general question posed in the Introduction the present results do not favour the view that the perception of speech is mediated by a different *type* of mechanism from that which operates upon other auditory stimuli. Indeed the fit of the independence models to data derived from words and phonemes in noise is, on the whole, as good as its fit to data derived from signals "tailored" to be applicable to the analysis.

However, in an important respect, the speech system *has* been shown to be different from that which deals with elementary stimuli. This difference has to do with the type of input to which the "analysers" respond in the two situations. In the word and the phoneme recognition studies we defined the underlying stimulus dimensions as phonemes (in the case of words) and the anatomical characteristics of the speech production apparatus (in the case of phonemes). The results apparently showed that these definitions held, thus α and β in Figure 1 have apparently been programmed to respond to the phonetic characteristics of the utterances. This feature has been noted in another context and has led to the "matching" or "active" theories of speech perception (Liberman, 1957; Liberman *et al.*, 1963; Stevens, 1960). These theories hold that there is a continuous implicit generation of speech by the listener which attempts to match his input and that a phoneme is identified when a correct match occurs. Such a process could effectively correspond to a set of passive analysers programmed to correspond to the *operations* of the listener as a talker.

Taken uncritically the results favour a description of the mechanism of word perception in terms of three banks of analysers. One bank may be assumed to consist of analysers which respond to complex acoustic *patterns*, characteristic of voicing, nasality, duration and affrication. The responses from this bank then feed to a second bank, programmed to respond to the patterns of inputs from the first bank characteristic of *phonemes*. The third bank would receive inputs from the second (phoneme) bank and respond with words. This description is both simplified and speculative, but would seem the most parsimonious explanation of the findings. It is, of course, possible that a non-orthogonal model would yield an equally good or better fit to the data, but such models are necessarily dependent upon the *degree* of correlation between dimensions. This value would require to be determined by experiment and would (if found to differ from zero) be specific to the parameter in question. The present paper has shown that the assumption of zero correlation between dimensions, coupled with the simplest decision mechanism *can* predict confusibility between words and phonemes within a restricted vocabulary.

This research was conducted at the U.S. Naval Electronics Laboratory, San Diego, California.

References

Broadbent, D. E. (1967). Word-frequency effect and response bias. *Psychol. Rev.*, **74**, 1–15.

Chaney, R. B., and Webster, J. C. (1965). Information in certain multi-dimensional acoustic signals. *U.S. Naval Electronics Lab. Rept. No.* 1339.

Conrad, R. (1964). Acoustic confusions in immediate memory. *Brit. J. Psychol.*, **55**, 75–84.

Corcoran, D. W. J. (1966). The prediction of responses to multidimensional stimuli from responses to unidimensional stimuli. *J. exp. Psychol.*, **71**, 47–54.

Corcoran, D. W. J. (1967). Perceptual independence and recognition of two-dimensional auditory stimuli. *J. acoust. Soc. Amer.*, **42**, 139–42.

Eriksen, E. W., and Hake, H. H. (1955). Multidimensional stimulus differences and accuracy of discrimination. *J. exp. Psychol.*, **50**, 153–60.

Green, D. M., and Swets, J. A. (1966). *Signal Detection Theory and Psychophysics.* London: Wiley.

Kimura, D. (1961*a*). Cerebral dominance and the perception of verbal stimuli. *Canad. J. Psychol.*, **15**, 166–171.

Kimura, D. (1961*b*). Some effects of temporal lobe damage on auditory perception. *Canad. J. Psychol.*, **15**, 156–65.

Kimura, D. (1964). Left-right differences in the perception of melodies. *Quart. J. exp. Psychol.*, **16**, 355–8.

Liberman, A. M. (1957). Some results of research on speech perception. *J. acoust. Soc. Amer.*, **29**, 117–23.

Liberman, A. M., Cooper, F. S., Harris, K. S., and MacNeilage, P. F. (1963). A motor theory of speech perception. *Proc. Speech Comm.*

Licklider, J. C. R., and Miller, G. A. (1951). The perception of speech. In Stevens, S. S. (Ed.). *Handbook of Experimental Psychology*, London: Chapman and Hall.

Miller, G. A., and Nicely, P. E. (1955). An analysis of perceptual confusions among some English consonants. *J. acoust. Soc. Amer.*, **27**, 338–52.

Milner, B. (1962). Laterality effects in audition. In Mountcastle, V. B. (Ed.). *Interhemispheric Relations and Cerebral Dominance*, Baltimore: Johns Hopkins Univ. Press.

Oldfield, R. C. (1966). Things, words and the brain. *Quart. J. exp. Psychol.*, **18**, 340–53.

Peterson, G. E., and Shoup, J. E. (1966). A physiological theory of phonetics. *J. Speech Hearing Res.*, **9**, 5–67.

Stevens, K. N. (1960). Towards a model for speech recognition. *J. acoust. Soc. Amer.*, **29**, 117–23.

Swets, J. A. (Ed.) 1964). *Signal Detection and Recognition by Human Observers.* London: Wiley.

Thorndike, E. L., and Lorge, I. (1944). *The Teacher's Wordbook of 30,000 Words.* New York: Columbia University Press.

Manuscript received 21st March, 1968.

AN INVESTIGATION INTO SOME OF THE UNDERLYING ASSOCIATIVE VERBAL PROCESSES OF THE STROOP COLOUR EFFECT

BY

DERRICK PRITCHATT

From the Department of Psychology, University of Leeds

The Stroop test was examined to throw some light on the verbal basis of the test. When words are presented printed in the wrong colour it is found that the written word provides more interference with response to the colour of print than occurs the other way round (the Stroop effect). It was shown that, by allowing subjects to match the colour of the print with colour patches instead of words, the interference due to the written word can be considerably reduced. In a further experiment, associations were formed by the subjects between nonsense syllables and colours in the course of a paired-associate learning task. When nonsense syllables were presented printed in colours different from those associated with them, it was found that the interference effects depended upon the direction of the stimulus-response association in the paired-associate experiment. It is concluded that the underlying processes in the Stroop effect are primarily verbal and depend partly upon the directionality of previously established verbal associations.

INTRODUCTION

When colour names are printed in the wrong colour they provide a striking interference effect. If, for example, the word RED is written in green print then the subject can be requested to report the written word and respond with RED or to report the colour of the print and respond with GREEN. Under these conditions it is found that subjects read the word fairly easily but experience considerable difficulty in reporting the colour of the print. Responses are given verbally and can be compared with the speed and accuracy of reading the names of colours in normal black print and also naming patches of colour since these two analogous conditions lack interference from simultaneously presented stimuli.

This technique was developed by Stroop (1935) and subsequently examined by various workers, e.g. Thurstone (1944) who used the test in a general perceptual study —unfortunately, he did not include the results in the factor analysis of the perceptual test battery. The colour names are usually presented in serial order and this has been done in the present study, but the effect is not strictly serial since it can be obtained by tachistoscopic exposure of individual items (Dalrymple-Alford and Budayr, 1966).

Most workers have concentrated either on the developmental aspects of the subjects' responses by investigating change as a function of chronological age (e.g. Ligon, 1932; Rand, Wapner, Werner and McFarland, 1963) or, more recently, the test has been used in conjunction with the study of drug effects (e.g. Callaway, 1959; Nash, 1962) though usually the intention has been to use the test to throw light on the action of the drug rather than the reverse. Much of this work has been summarized in a most useful recent review by Jensen and Rohwer (1966).

Not a great deal of work has been done which throws light on the underlying processes involved in the test (a notable exception being that of Klein, 1964, which will be discussed later) and the present study was undertaken in the hope of pursuing this matter further. Three experiments are reported:

(1) A repeat of the original Stroop test but with the conditions adapted to those used throughout the present study in order to obtain some comparison of the magnitude of the effects.

(2) A modification to investigate the importance of the verbal component of the effect.

(3) An attempt to manipulate the effect by means of learned interference.

EXPERIMENT 1

Subjects

Subjects were 14 male and 9 female university undergraduates between the ages of 18 and 35. Each subject was used in every experimental condition. Before the experiment, each subject was given the Ishihara Colour Test but only one subject with a red-green defect had to be replaced. One other subject was replaced later—see discussion below.

Materials

All words were printed in Standard Ostwald colours on 4 in. × 6 in. cards. The colours used were: green, red, yellow, blue, orange, purple. All words were printed with a stencil and there were 6 words per card. Colour patches used were made the same height and width as the longest words, the same number of items was used as for words. Six cards at a time were displayed in a single vertical column by means of a holder in which their places could be quickly interchanged. It was arranged that each stimulus appeared once on a card in each of the positions 1 to 6, also each card appeared once in each of the positions 1 to 6.

Where colour names were printed in the wrong colour, the following near-complementary colours were always taken: red and green, blue and orange, purple and yellow. All times taken were recorded to 0·1 sec. using a stopwatch.

Procedure

The six conditions used are summarized below:

Condition code	*Stimulus*	*Response required*
P	Colour patches.	Colour names.
B	Colour names, black print.	Read word.
WW	Colour names, printed wrong colour.	Read word.
WC	Colour names, printed wrong colour.	Colour of print.
RW	Colour names, printed right colour.	Read word.
RC	Colour names, printed right colour.	Colour of print.

Since it was impossible with 22 subjects to use every possible combination of conditions an approximate counterbalancing procedure was used in which each condition occurred equally often at the beginning, middle and end of the experiment. Also, each condition was preceded by a different condition each time it appeared. Finally, two trials were run for every condition (one immediately after the other) and to further minimize position effects only the second trial was used in the final analysis; the results obtained being very similar to those from the first trial except for the fact that the overall variance was reduced.

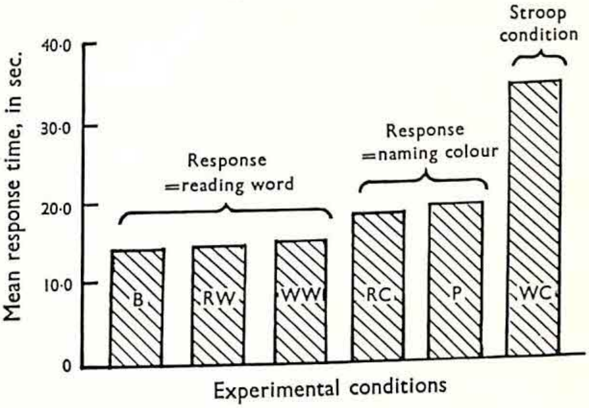
Subjects were seated 6 ft. in front of the empty display board and were asked to fixate approximately on the middle of the board and close their eyes. The room was dark and the board was lit by two shielded photoflood lamps positioned each side of the board.

Standard instructions were used and subjects were given a list of colours to be used so that they did not have difficulty in naming, e.g. purple (this had happened on trial runs). In order to obtain maximal interference effects, subjects were told to read down as quickly as possible but to correct any errors as they went along; thus, errors and time acted in the same direction. Two trials were given for each of the six conditions and between each trial the subject closed his eyes for 1 min. In this time, the experimenter changed the order of the cards and, after the second trial, the condition. The time taken by the subject to respond to all 36 stimuli was noted as well as any errors made.

Results

Because of the instruction to correct errors, relatively very few were made and will not be further discussed. The results for the main dependent variable, time, are shown in Figure 1.

FIGURE 1



The mean reaction time for each of the experimental groups, P, B, WW, WC, RW and RC (for explanation of symbols, see text).

Comparison between sexes for each condition yields non-significance ($p > 0{\cdot}05$ in every case, using Mann-Whitney *U* two-tailed test), the data from different sexes was, therefore, pooled in subsequent comparisons. In the case of naming colour patches, Stroop (1935) had found females to be faster but in the present experiment this tendency (male mean = 19·4 sec., female mean = 18·5 sec.) did not reach significance even for a one-tailed test.

Bartlett's test for homogeneity of variance was carried out on the data and there was found to be a significant lack of homogeneity ($p < 0{\cdot}01$). A Friedman analysis of variance was, therefore, used to examine differences between means of experimental conditions, and there was found to be an overall significant difference ($p < 0{\cdot}001$). Differences between adjacent conditions was tested by means of the Wilcoxon test and the following results were obtained: $\underline{\text{B RW WW}}$ $\underline{\text{RC P}}$ WC where, adjacent conditions underlined do not differ significantly ($p > 0{\cdot}05$), the gaps being significant at $p < 0{\cdot}01$ (two-tailed test).

It can be seen that the data fall into three groups: reading words, naming colours and the Stroop condition. When words are printed in the wrong colour, the subject finds it easier to respond to the word and ignore the colour of print than he does to respond to the colour of print and ignore the word. The strength of this Stroop effect is very marked: one subject with a speech defect, previously unsuspected by the experimenter, began to stutter so badly in condition WC that he was unable to continue and was replaced.

It is interesting to note that the fact that the word is printed in the correct colour (a condition not tested by Stroop) did not facilitate reading compared with the black print condition—presumably because of the familiarity of black print.

EXPERIMENT 2

In order to study the extent of the verbalizing component an attempt was made to minimize it by allowing the subject to match the colour of the print with colour patches. Of course, there is no guarantee that he does not verbalize at all but the degree of success in verbal suppression as distinct from the new motor components of the task can be assessed to some extent by re-introducing the verbal component in the response by making the subject match against written words.

Subjects

Ten male and 10 female students, between the ages of 18 and 30 (age and sex was equated between groups) were to have served as subjects; however, at the last minute two subjects (one male and one female) withdrew from one group.

Materials

The stimuli were the same as for Experiment 1 except that words printed in the correct colour were now omitted. Instead of verbal report, the subject responded on a set of keys laid out in front of him. On top of each key was a display card which could be changed. The cards bore either colour patches or the names of colours, so that the subject could respond by selecting either the colour patch or the name of the colour printed in black, depending upon the conditions given below.

Procedure

Two experimental groups were run, these groups responded either on patch keys or word keys. The order of the display cards on the keys was constant for any one subject but varied between subjects so that each colour patch or name appeared at least once on each key. There were four tasks to be completed for each group and these tasks were counterbalanced as far as possible—particular care was taken with representation in the first and last conditions. The eight conditions used are summarized below:

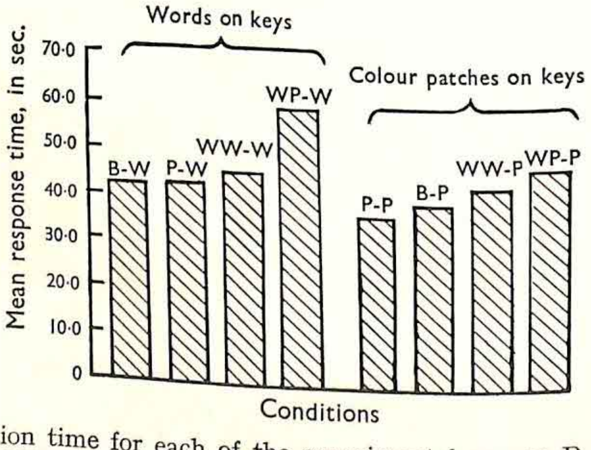
Condition code	*Stimulus*	*Response required*	*Response key*
P–W	Colour patches.	Colour.	Word.
B–W	Colour name, black print.	Read word.	Word.
WW–W	Colour name, printed wrong colour.	Read word.	Word.
WP–W	Colour name, printed wrong colour.	Colour of print.	Word.
P–P	Colour patch.	Colour.	Patch.
B–P	Colour name, black print.	Read word.	Patch.
WW–P	Colour name, printed wrong colour.	Read word.	Patch.
WP–P	Colour name, printed wrong colour.	Colour of print.	Patch.

All procedures such as trials, inter-trial times, etc., were as in Experiment 1 except that subjects were requested to make their response by pressing the appropriate key. The times taken by each subject to complete the tasks were noted by means of a stopwatch.

Results

The results are shown in Figure 2.

FIGURE 2



The mean reaction time for each of the experimental groups P–W, B–W, WW–W, WP–W, P–P, B–P, WW–P and WP–P (for explanation of symbols, see text).

Bartlett's test applied to the data showed no significant departure from homogeneity of variance ($p > 0{\cdot}05$). A one way analysis of variance for repeated measurements was carried out with the results shown in Table I.

TABLE I

Source	*Sums of squares*	*d.f.*	*Variance*	*F*	*p*
	Words on keys				
Between people . .	1175·7	7			
Within people . .	2137·3	24			
Conditions	1559·9	3	520·0	18·9	0·001
Residual	577·4	21	27·5		
Total	3313·0	31			
	Colour patches on keys				
Between people . .	2009·2	9			
Within people . .	933·2	30			
Conditions	604·3	3	201·4	16·5	0·001
Residual	328·9	27	12·2		
Total	2942·4	39			

An overall difference between experimental conditions can be seen both for colour patches on keys and for words on keys (in each case, $p < 0{\cdot}001$). The Newman–Keuls test was used to examine all gaps between conditions with the following results (where adjacent conditions underlined do not differ significantly, $p > 0{\cdot}05$).

Colour patches on keys = $\underline{\text{P–P}\quad\text{B–P}}\qquad\underline{\text{WW–P}\quad\text{WP–P}}$
(significance of gap, $p < 0{\cdot}05$)

Words on keys = $\underline{\text{B–W}\quad\text{P–W}\quad\text{WW–W}}\quad\text{WP–W}$
(significance of gap, $p < 0{\cdot}01$)

It can be seen that the introduction of patches as a response in the Stroop condition very much reduces the interference effect relative to the adjacent condition and that the interference re-appears with the introduction of words.

EXPERIMENT 3

Experiment 2 shows that the order of operations (e.g. responding with a word or to a word) might be fairly critical in the interference effect. To examine this further, an attempt was made to accustom the subjects to carrying out different orders of the same operation and testing the consequent interference in a Stroop simulated task.

Subjects

Two experimental groups, each consisting of eight subjects, were used. Each group contained the same number of males and females, and the mean age of each group was very nearly the same.

Procedure

Subjects in both groups were first required to learn a list of 3-letter nonsense syllables matched for Low association value (taken from Hilgard, 1951). The nonsense syllables were presented one at a time (for 2 sec. with 3 sec. intervals) on the table and the subject attempted to recall them in any order. Following this, the nonsense syllables were again presented, in a different order, and the subject again attempted to recall. This was repeated to a criterion of three consecutive runs all correctly recalled. The number of trials to criterion was taken into account in the matching of the experimental groups though, in fact, the variation between subjects was very small. Having familiarized themselves with the nonsense syllables, both groups of subjects now learned to associate a colour with each of these nonsense syllables. Times of display and intertrial times were as previously and subjects were run to a criterion of three trials without error. Nonsense syllables and colour names were presented as in an ordinary paired-associate experiment and were learned by the anticipation method. The critical difference between the two experimental groups was that one group always received a nonsense syllable as the first member of the pair and responded with the name of a colour whereas the other group always received a colour as the first member of the pair and responded with a nonsense syllable. It was thus hoped to establish some directionality of association thus: nonsense syllable ——→ colour and colour ——→ nonsense syllable.

In the final stage of the experiment, subjects were presented with a list of nonsense syllables printed in a different colour from the one which they had learned as a pair. They were asked to respond to the nonsense syllable with the colour they had learned, and to the colour of the print with the nonsense syllable they had learned. All responses were given verbally as in Experiment 1. They were also presented with colour patches and black nonsense syllables as stimuli for comparison. The conditions used are summarized below:

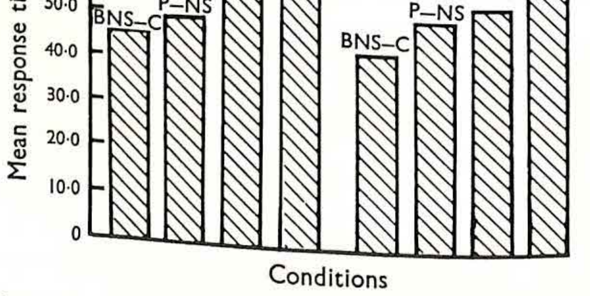
Condition code	*Stimulus*	*Response required*
BNS–C	Nonsense syllables, black print.	Colour.
P–NS	Colour patch.	Nonsense syllable.
WNS–C	Nonsense syllable, wrong colour.	Nonsense syllable associated with colour.
WNS–NS	Nonsense syllable, wrong colour.	Colour associated with nonsense syllable

As previously, the dependent variable was time taken to complete each condition, measured by means of a stopwatch.

Results

The results are shown in Figure 3.

FIGURE 3



The mean reaction time for each of the experimental groups BNS–C, P–NS, WNS–C and WNS–NS (for explanation of symbols, see text).

Bartlett's test again showed no significant departure from homogeneity of variance ($p > 0{\cdot}05$). A one way analysis of variance for repeated measurements gave the results shown in Table II.

TABLE II

Source	*Sums of squares*		*d.f.*		*Variance*	*F*	*p*
Train: stimulus = colour, response = nonsense syllable							
Between people		8285·0		7			
Within people		7390·5		24			
Conditions	3016·8		3		1005·6	4·8	0·05
Residual	4373·7		21		208·3		
Total		15675·5		31			
Train: stimulus = nonsense syllable, response = colour							
Between people		2144·5		7			
Within people		2824·2		24			
Conditions	1306·1		3		435·4	6·0	0·01
Residual	1518·1		21		72·3		
Total		4968·7		31			

In each case there was an overall significant difference between conditions: for train colours, response nonsense syllables ($0{\cdot}05 > p > 0{\cdot}01$) and for train nonsense syllables, response colours ($p < 0{\cdot}01$). The Newman-Keuls test was used to investigate where these differences lay and the following gaps were found to be significant.

For train colour, with response nonsense syllable:

BNS–C and WNS–C ($p < 0{\cdot}05$)
BNS–C and WNS–NS ($p < 0{\cdot}01$)
P–NS and WNS–NS ($p < 0{\cdot}05$)

For train nonsense syllable, with response colour:

BNS–C and WNS–NS ($p < 0{\cdot}05$)
BNS–C and WNS–C ($p < 0{\cdot}01$)
P–NS and WNS–C ($p < 0{\cdot}05$)

The picture is more complicated than previously but it can be seen that the direction of the learned association has affected the Stroop interference. A subject who has learned to respond to the nonsense syllable and is now required to respond to the colour of print, finds the written nonsense syllable the source of greatest interference (i.e. WNS–C is the most difficult condition). However, a subject who has learned to respond to the colour and is now required to respond to the nonsense syllable finds the colour of print the source of greatest interference (i.e. WNS–NS tends to replace WNS–C as the most difficult condition—a kind of "inverted Stroop").

Discussion

From Experiment 1, it can be seen that when words printed in the wrong colour are read there is little or no interference due to the incongruous colour of the print. However, when the colour of the print is named there is considerable interference from the written word. This suggests that the effect is a purely verbal one, i.e. when the colour of the print is to be verbalized there is a conflict with the existing word during the processing of the verbal label of the print colour. Certainly this kind of approach has been the basis of those developmental studies which have looked for a change of effect with increasing verbal sophistication due to age; but this evidence is very indirect. The only direct test carried out previously is that of Klein (1964) who used various words printed in colour and found that the extent to which the written word interferes varies directly with the relevance to colour of the written word and more generally to its familiarity when words of different English usage are compared (this work has been confirmed in every detail by the present writer who ran all of Klein's experiments some years ago before he was aware that they had been carried out). The other direct test is provided by Experiment 2 which shows that when the verbal aspect of the colour print is minimized in the patch key condition, the Stroop effect is much less but the effect returns when words are put on the keys. It is interesting to note that even in the case of matching colour print to colour patches on the keys (WP–P) there is some interference from the written word but, of course, it is possible that the subject verbalizes mentally when carrying out the task.

A "complete" explanation of the Stroop effect undoubtedly awaits much more empirical evidence. What will be attempted below is simply to show that some types of response involved in stimulus identification, being apparently more difficult to perform, are more susceptible to interference of the Stroop type than others.

Consider the following stimulus response processes involving verbal attributes:

object → verbal label
verbal label → object
verbal label → verbal label

Of these, verbal label → verbal label is clearly easier than object → verbal label (B is faster than P, Experiment 1) and the latter would be expected to be more susceptible to interference irrespective of the nature of the interfering stimuli (this does not seem to have been tried by any investigator). The difference between these conditions extends beyond colour names to the naming of photograph of common objects compared with the reading of the names of these objects (unpublished data from this laboratory). Since object → verbal label is required in the Stroop condition WC but not in WW it seems likely that it is a contributory factor to the Stroop effect.

Object → verbal label is also more difficult than verbal label → object (Experiment 2). It is interesting to note that Gazzaniga and Sperry (1967) have shown that a subject with a sectioned corpus callosum which isolates the speech hemisphere of the brain can readily identify, say, a spoon by touch from an assembly of objects they cannot view directly when supplied with the verbal label "spoon" to the right hemisphere even though he is unable to then produce the verbal label.

Presumably the ease with which verbal label → object can be carried out is the result of the heavy emphasis placed on reading in our society; whereas object → verbal label is not always required. We do not verbalize all objects and, indeed, quite complicated tasks like concept formation can be carried out without the subject necessarily being able to verbalize in perfect step with his sorting into categories (Heidbreder and Overstreet, 1948). That the direction of a learned association is, indeed, important in interference effects is shown clearly by Experiment 3.

There is the possibility that lack of appreciable interference in the WW condition

(Experiment 1) occurs because the coloured print can be ignored totally and is never processed as sensory information. This hypothesis predicts that in the WC (Stroop) condition both stimuli reach short term store, whereas in condition WW only one stimulus is ever actually received. Preliminary trials have been run in this laboratory in which subjects view six names of months printed in different colours presented tachistoscopically. The months and colours were matched for length and informational content as both came from a limited response pool known to the subject—this is important if a more indirect method of investigation is to be avoided (e.g. that of Grand and Segal, 1966). Subjects viewed the stimuli and were requested to read, as quickly as possible, either the words or the colour of the print, the viewing time being adjusted so that the subject could barely complete the task. After they had done this, they were asked to recall the stimuli (i.e. words or colour of print) to which they had *not* been asked to attend. The data collected so far show that in either condition, subjects do allow both stimuli to reach short term storage since some correct information can be given about both stimuli. This approach is being pursued to see if there is a difference in the amount of "irrelevant" information that can be reported in each of these conditions and also to investigate what happens when the list is long enough to exceed the short term memory for earlier items since the time that an interfering stimulus is in short term store can be very important (Posner and Konick, 1966).

REFERENCES

CALLAWAY, E. (1959). The influence of amobarbitol (amylobarbitone) and methamphetamine on the focus of attention. *J. Ment. Sci.*, **105**, 382–92.

DALRYMPLE-ALFORD, E. C., and BUDAYR, B. (1966). Examination of some aspects of the Stroop colour-word test. *Percept. mot. Skills*, **23**, 1211–24.

GAZZANIGA, M. S., and SPERRY, R. W. (1967). Language after section of the cerebral commissures. *Brain*, **90**, 131–48.

GRAND, S., and SEGAL, S. J. (1966). Recovery in the absence of recall: an investigation of colour-word interference. *J. exp. Psychol.*, **72**, 138–44.

HEIDBREDER, E., and OVERSTREET, P. (1948). The attainment of concepts: V. Critical features and contexts. *J. Psychol.*, **26**, 45–69.

HILGARD, E. R. (1951). Methods and procedures in the study of learning. In S. S. STEVENS (Ed.), *Handbook of Experimental Psychology*. New York: Wiley.

JENSEN, A. R., and ROHWER, W. D. (1966). The Stroop colour-word test: a review. *Acta Psychol.*, **25**, 36–93.

KLEIN, G. S. (1964). Semantic power measured through the interference of words with colour-naming. *Amer. J. Psychol.*, **77**, 576–88.

LIGON, E. M. (1932). Genetic study of colour naming and word reading. *Amer. J. Psychol.*, **44**, 103–22.

NASH, H. (1962). *Alcohol and Caffeine*. Illinois: Thomas.

POSNER, I. M., and KONICK, A. F. (1966). On the role of interference in short term memory. *J. exp. Psychol.*, **72**, 221–31.

RAND, G., WAPNER, S., WERNER, H., and MCFARLAND, J. M. (1963). Age differences in performance on the Stroop colour-word test. *J. Personal.*, **31**, 534–58.

STROOP, J. R. (1935). Studies of interference in serial verbal reactions. *J. exp. Psychol.*, **18**, 643–62.

THURSTONE, L. L. (1944). *A Factorial Study of Perception*. Chicago: Univ. Chicago Press.

Manuscript received 25th May, 1968.

TYPING PERFORMANCE AS A FUNCTION OF TEXT

BY

L. H. SHAFFER and J. HARDWICK

From the Department of Psychology, University of Exeter

The performance of skilled typists was examined on texts ranging from prose to random letters, with the aim of discovering some of the factors that support its fluency. It was found that speed and errors were the same for prose and random word texts, and although performance got progressively worse on more degraded texts the biggest decline came between first- and zero-order random letter texts. The results were discussed in the context of several hypotheses about transcription processes, some assuming a statistical, left-to-right encoding of letters and others assuming linguistic encoding, either preserving the linguistic units in response units or translating the linguistic units into letter responses under the control of a sequencing process.

INTRODUCTION

What has the skilled typist learned in mastering typewriting? Several experiments were carried out to obtain information relevant to this problem by examining the speed and errors of professional touch typists on a series of texts ranging between prose and random letter strings.

The study is motivated by the following observations: a moderately good typist can type prose at a rate at which inter-response time is less than simple reaction time observed in discrete trials, yet however she may encode the stimulus input the fact remains that response output is produced letter by letter.

The following studies serve as background. Butsch (1932) found that the eye-finger span in typing is a random variable that might range, within the performance of a subject, between 0 and 13 letters. For different subjects at different levels of proficiency the average span per subject varied between 3·2 and 7·6 letters, and there was a correlation between rated speed (he did not measure speed in his tests) and average span. The world champion typist was tested on a variety of texts, presumed to differ in difficulty, and his average span ranged between 5·6 and 7·7 letters. Butsch observed that the average eye-finger span typically covers an amount of material that can be transcribed in about 1 sec.

Hershman and Hillix (1965) used an ingenious method of mounting the text on the carriage of the typewriter so that every time the subject struck a key the text moved along one symbol past an aperture in a screen. Exposing 1, 2, 3 or 6 letters at a time or providing unrestricted view they found a negatively accelerated increase in typing speed with amount of exposure. Also, using texts of prose, random words or random letters—the latter texts being derived from the first—they obtained higher speeds with the less random texts, the gain in speed increasing with level of exposure.

The better typists could type prose some five or six times faster with unrestricted view than with exposure of one letter at a time. Clearly finger dexterity and over-learned associations between letters and finger movements play only a small part in the typist's skill, the rest lies in her ability to read ahead and to exploit structure (perhaps accidental) in the text. Even with texts having only the letter frequency of English she can more than double her speed by being allowed to read ahead. Quastler and Wulff (1965) obtained transmission rates of 17 bits per sec. with random letter texts and unrestricted view.

A study by Genest (1956) is interesting for its recording technique. A motor drove the roller of the typewriter at a slow fixed speed and so provided a cumulative record of times between each key press. He was thus able to observe that pauses tended to occur at points of linguistic difficulty, and that with degraded texts there was a loss of both typing speed and regularity. His analysis of errors (Genest, 1957) is not always convincing, but it showed fairly clearly that errors occurred more often with degraded texts than with prose and were more likely to implicate the less frequently used letters of the alphabet. A high proportion of errors arose from pressing a key adjacent to the correct one.

A word about the construction of texts. Shannon (1948) showed that one can construct successive approximations to prose by imposing statistical constraint over longer and longer sequences in letter or word strings. This method has been used to study serial performance (e.g., Sumby and Pollack, 1954) but is not adopted systematically here for the following reasons: (*a*) The method, being statistical, is susceptible to sample size and guarantees a representation of a given level of structure (and discriminability between different levels) only in indefinitely long strings. This dependence on sample size becomes more acute with higher order approximations. (*b*) It seems natural to analyse language into such units as syllables, words and phrases, and it is likely that an increase in statistical constraint on letter strings merely increases the probability of reconstructing these units as artefacts. In this case it is preferable to introduce the required units directly than to rely upon their chance occurrence. This should not prejudice any theoretical issue as long as the alternative method of producing text does not give abnormal distributions of letter sequences of length $k = 2,3\ \ldots$ (k-grams).

EXPERIMENT I

Transcription was examined on 10 passages of text, each about 250 letters long, ranging between prose and random letter strings. The experiment provides a test of two conjectures: (1) that typing speed increases with the frequency that syllable-, word-, or phrase-like segments appear in the text and is progressive with the size of linguistic unit; and (2) that speed is limited by the task of reading letters in the interior of long words (a word is defined as a string of letters uninterrupted by a space character) and that this factor interacts with text structure.

Method

Subjects. These were all qualified touch typists employed in the University. A criterion was set in advance that anyone making more than 2½ per cent. errors in transcription would be rejected from the group whose times were analysed and many girls were in fact rejected on this basis. Their errors were, however, analysed and their times are still available. In the final selection there were 20 girls, 10 in each of two groups. They were tested separately.

Apparatus. Each girl used her own typewriter. Most machines were Olivetti 82's (mechanical) and the others mainly IBM Executive's (electrical). Each text was printed completely in lower case letters in 10-point type, with 1½ spacing between lines. The time taken to transcribe a passage was recorded to 0·1 sec. with a Venner timer.

Text. The 10 passages were as follows:

Prose, P_E and P_D: an easy and a difficult passage of prose, P_E taken from an article on gardening and P_D from a text on linguistics. Ease or difficulty is only vaguely defined, as a first approximation, in terms of content and word familiarity. The passages were edited to have little punctuation and no names or digits.

Word, W_E and W_D: random rearrangements of the words in P_E and P_D respectively.

Syllabic, S_F and S_R: obtained from P_D by shuffling the spaces between words into syllable boundaries within words. The new words thus created were taken in reverse order to give S_F, having words about six letters long, and S_R, having word lengths distributed as in P_D. (Suffixes F and R stand for fixed and random length.)

First order, L_{1F} and L_{1R}: random letter strings having the distribution of letters in P_D. The words in L_{1F} were six letters long, while word lengths in L_{1R} were distributed as in P_D.

Zero order, L_{0F} and L_{0R}: random letter strings in which all letters were equiprobable. Word lengths matched those in L_{1F} and L_{1R} respectively.

Design. The first 10 typists were given the 10 passages in random order, using different orders for each typist according to a latin square. Several girls commented afterwards that having typed nonsense passages initially they approached all subsequent passages as nonsense. To check on such an effect the next 10 typists were given the passages in invariant order going progressively from prose to zero order text.

Instruction. The typists were asked to transcribe the passages exactly, as quickly as they could without error. If errors were made they were not to correct them and should avoid stopping in midpassage. It was redundantly stressed that they were not to introduce upper case letters or punctuation or to alter line layout.

Each sheet of text was placed in a reading position to suit the typists and they were told to look at it and start typing as soon as they heard the (loud) click of the switch that started the timer. On typing the terminal full stop in a passage they were to call "stop." They were given practice with five brief passages to get used to typing under these conditions.

Results

Time scores. The times per passage were converted to mean time per symbol, counting all letters, spaces and punctuation marks as symbols. Carriage return time was estimated at about $\frac{3}{4}$ sec. for mechanical machines and about $\frac{1}{2}$ sec. for electrical ones. Thus mean time per symbol was estimated by subtracting carriage return time from the total time and dividing the remainder by the number of symbols.

The average response times on each passage, for the 10 subjects in each group, are shown in Table I. Subjects in group A received the passages in random order and those in group B in fixed order. It is immediately clear that there were only minor differences in speed between easy and difficult prose and trivial differences between prose and random words, or between passages with nonsense words of fixed and random lengths.

TABLE I

MEAN RESPONSE TIMES PER SYMBOL IN MSEC. IN EACH OF THE 10 PASSAGES. EXPERIMENT I

		Prose P	*Word* W		*Syllabic,* S	*First order,* L_1	*Zero order,* L_0
Group A	Easy	151	152	Fixed	195	239	329
	Difficult	157	155	Random	197	239	332
Group B	Easy	166	163	Fixed	222	265	357
	Difficult	171	167	Random	217	268	359

Analyses of variance were carried out on the results of each group. In group A the subjects variable was significant, F (9,72) = 11·3, $p < 0{\cdot}001$; the passages variable was significant, F (9, 72) = 232, $p < 0{\cdot}001$, but not the order in which they were typed, F (9, 72) = 1·49. In group B the subjects variable was significant, F (9, 81) = 51, $p < 0{\cdot}001$: the passages variable (now confounded with order of typing) was significant, F (9, 81) = 98, $p < 0{\cdot}001$.

Further post hoc tests within the passages variable, using the Scheffé method of contrasts (Scheffé, 1959), showed that there were significant differences at the $p = 0{\cdot}001$ level of significance, for both groups, between the following sets of passages: P and W together and S; S and L_1; L_1 and L_0.

Response times were slightly longer in group B than in group A, but this is an accident of sampling. Because sampling was non-random the difference could not in any case be tested statistically. For the same reason we could not test whether the different methods of presentation between the two groups had any effect. However the predicted effect is that the difference between speeds with prose and random texts should be smaller in group A than in group B and Table I shows that there is little evidence of this.

Error scores. Any textual discrepancy between original and transcript was counted as an error and, for the purpose of scoring, the unit of error was a symbol—letter, space or punctuation mark.

TABLE II

DISTRIBUTION OF ERRORS FOR EACH KIND OF TEXT. EXPERIMENT 1

		Prose, P	*Word, W*	*Syllabic, S*	*First order, L_1*	*Zero order, L_0*	*Total*
Groups A and B	Omission	19	11	24	15	34	103
	Response	19	31	27	32	108	217
	Reading	3	2	8	14	20	47
	Context	19	19	34	56	46	174
	Random	3	2	4	13	22	44
	Total	63	65	97	130	230	585
Rejected subjects	Omission	21	23	36	46	82	208
	Response	25	38	53	43	113	272
	Reading	2	0	15	20	41	78
	Context	27	17	30	51	40	165
	Random	5	6	11	15	41	78
	Total	80	84	145	175	317	801

Subjects in groups A and B made altogether 585 errors, which represents an error rate of 1·2 per cent. Their distribution over the main kinds of passage is shown as the column totals in Table II. They show a monotonic increase in degraded texts, the greatest increase being in L_0.

Any attempt to classify errors by their causes is hazardous because most errors are cryptic and their possible causes ambiguous, sometimes in several ways. However a classification was developed, based initially on 13 classes, many of them similar to those used by MacNeilage (1964), that were subsequently collapsed onto five, which we label Omission, Response, Reading, Context and Random. In case of ambiguity an arbitrary priority of assignment was given to Response, then to Reading and then to Context classes. As an exception to this priority any transposed pair of adjacent letters was classed as a Context error.

Response errors were those made by hitting a key adjacent to the correct one, which could be any key, or the space bar, in the hexagonal field of the correct key. Reading errors were those letters that are confusable, visually or acoustically (cf. Conrad, 1964), with the correct one. Context errors were transpositions of adjacent letters and displacements of letters appearing within a range of three letters left or right of the error position, also a few cases in which the error made a nonsense word more like a dictionary word.

Table II shows the class by text breakdown of errors. There is a significant contingent relationship between class of error and text, χ^2 $(d.f. = 16) = 55{\cdot}1$, $p < 0{\cdot}001$, a major part of which is due to a relatively large number of Context errors in L_1 and a decline of these in L_0.

What of the errors made by the reject subjects? They made altogether 801 errors and these are classified in the bottom half of Table II. The pattern is reasonably similar to that in the upper half, thus giving some support to the method of classification.

Experiment 2

A surprising result of the first experiment was that random words were typed as quickly as prose. Both Hershman and Hillix (1965) and Genest (1956) say that prose was typed the faster, although neither gave statistical support for this. It seemed worth repeating this part of the study using different texts.

Method

Subjects. Ten girls were tested who had served in the first experiment.

Text. Five passages of prose were used, each 250 letters long and edited as before. Three passages were taken from novels by Hemingway, the other two were from texts on Geography and Physics. Whereas the Hemingway passages were conceptually simple and used only short words the textbook passages were conceptually difficult and contained many long, relatively unfamiliar words. Five other passages were produced by randomizing word order of the prose passages.

Design. The 10 subjects were given each of the 10 passages in a different order using a latin square. All further details were the same as in Experiment 1.

Results

Since there were altogether 94 errors, which represents an error rate less than 0·4 per cent., the analysis of results is based upon time scores only. The times per passage were converted to mean times per symbol, as before, and these means are shown in Table III.

It is clear once again that while random word passages, W, take slightly longer than prose passages, P, at least in easy texts, the difference is trivial. There is however a rather more clearcut difference between easy (E) and difficult (D) texts than was formerly obtained.

TABLE III

Mean Response Times per Symbol in MSec. Experiment 2

	Easy			*Difficult*	
	E_1	E_2	E_3	D_1	D_2
Prose	140	139	141	152	149
Word	143	140	142	155	148

An analysis of variance on these data shows that the text variable is significant, $F(9, 72) = 12{\cdot}2$, $p < 0{\cdot}001$. The use of the Scheffé post hoc contrasts test shows that the difference between E and D texts is significant at the level $p = 0{\cdot}001$, and this seems to be the only contrast of practical interest that is significant.

Since there is no difference between P and W passages it is adequate, for transcription, to define text difficulty in terms of word difficulty alone.

EXPERIMENT 3

The first experiment showed that the step in degrading text that most affected performance was between first- and zero-order letter texts (L_1 and L_0). This might be due to a decline in syllable-like artefacts, because of fewer vowels, or to the relative increase in appearance of the less frequently used letters of the alphabet (Genest, 1956). An attempt was made to separate these factors by maintaining equal frequencies of all letters of the alphabet but manipulating letter order in different texts.

Also a further test was made for an effect of reading letters within long words. The first attempt in Experiment 1 was weak in that it compared short against random length words. It is preferable to test short against long words.

Method

Subjects. Four girls were tested who had served in the first experiment.

Text. Four passages of text were constructed each using every letter of the alphabet 10 times. One passage selected the 260 letters at random and displayed these in words five letters long. Another passage selected these letters five at a time such that as far as possible they could be made into readable syllables, e.g. . .spowd. ., . .throx. . . Two further texts were obtained by grouping three words in each of these passages into a single word, so that all words had 15 letters, except for a terminal word of 20 letters.

Design. The four subjects typed each of the four passages in different sequence using a latin square.

Results

The left half of Table IV shows mean response times per symbol in each condition. Clearly, both word structure and word length are effective variables and their effects are additive. There is no need to test these results statistically because they appear almost invariant in each subject.

TABLE IV

MEAN RESPONSE TIMES PER SYMBOL IN MSEC. EXPERIMENT 3

			(From Experiment 1)		
	Syllable	*Random*		*First order*, L_1	*Zero order*, L_0
5-letter	246	326	Fixed	236	344
15-letter	292	373	Random	242	343

The right half of Table IV shows mean times for the same subjects with comparable passages, L_1 and L_0, in the first experiment. For the random letter texts the 5-letter words were typed slightly faster than the 6-letter words in L_0. This could easily be an effect of letter sampling in the two texts but may also reflect word length. On the other hand the syllabic words were typed only slightly slower than first order text.

DISCUSSION

The simplest hypothesis about transcription, since typing is performed letter by letter, is that text is encoded also in this way. If one supposes that typing speed is a function of the likelihood of the letters typed and that this likelihood is conditional upon the sequence of letters that preceded a given letter, then this leads to the notion of a left-to-right finite state computer (Chomsky, 1963).

Such a computer has distinct states for all letter sequences of any given length (k) and in each state a probability distribution over the set of states that can come next. The transition from one state to another leads to the transcription of a symbol of text. For example, if $k = 3$ and the computer is in state ..num.. then with probability p(num; b) it will have to transmit the letter ..b.. and go to state ..umb.., while with probability p(num; e) it will have to transmit ..e.. and go to state ..ume.., and so on. Note that k is the effective memory span of the computer. For a set of 27 symbols—26 letters of the alphabet and a space—the uncertainty per symbol, if they are equally likely to occur, is 4·75 bits. If it is known that the symbols occur as English text then Shannon (1951) has estimated that the uncertainty is 4·03 bits per symbol if $k = 0$ (zero-memory), 3·32 if $k = 1$, 3·1 if $k = 2$, and as k increases to cover more and more of the previous text the uncertainty declines to between 0·6 and 1·3 bits per symbol.

If Shannon's estimates are even approximately correct then it is seen that the relationship between text uncertainty and typing speed is monotonic but not linear. As Table I shows, about half the loss in speed between prose and random letters occurred between L_1 and L_0, but this last step accounts for only a fraction (0·72 bits) of the overall increase in text uncertainty.

Random words were typed as fast as prose, so that k-gram analysis would not extend effectively beyond word boundaries, while random words were typed faster than syllable strings, so that k would have to be an integer large enough to permit discrimination between them. If the average length of words in all texts was around five or six letters then these results impose incompatible requirements on k. Consider a k-gram scan of the word ..machine.. when $k = 3$, 4, 5 or 6:

--m	---m	----m	-----m
-ma	--ma	---ma	----ma
mac	-mac	--mac	---mac
ach	mach	-mach	--mach
chi	achi	machi	-machi
hin	chin	achin	machin
ine	hine	chine	achine
ne-	ine-	hine-	chine-

As k increases the k-gram analysis is underemployed in more and more of the early steps while, on the other hand, overall there are few steps in which the continuation on the right is highly determined. Thus there seems to be no value of k that would permit discrimination corresponding to the speed differences between prose and syllabic text such as ..tution substi softhe interm...

Alternatively, by increasing k and modifying the starting conditions one turns this computer into a device that encodes the spelling of proper words as a branching tree developing from left to right. Starting with a space symbol the device goes to one of the 26 first letters of words. From this it takes one of a set of branches going to second letters of words, and so on. Each left to right path through the tree spells out a word and terminates in a space symbol (which guarantees a divergent branch between words if the spelling of one word is embedded in that of another, e.g., ..ever.., ..every..). Such a device would reject nonsense words because sooner or later it would fail to produce a continuation on the right, and beyond this point all letters would be treated as random. Hence the device would show poor discrimination between such texts, for instance, as L_1 and L_0, contrary to the results.

Given Shannon's estimates of residual uncertainty in texts as redundant as prose it is clear that the speeds attained by typists are still greater than one would expect from laboratory results on choice reaction time.

So far we have withheld comment on the fact that typists read ahead of their typing, whereas in finite state transcription the information that makes a letter more likely, and so easier to recognize and transmit, is contained on the left. A finite state hypothesis would fail, after all, to account for the fact that typing speed is highly dependent upon being able to read ahead in the text (Hershman and Hillix, 1965).

Suppose instead that, in reading, text is analysed into such linguistic units as syllables, words and short phrases such as ..a book.., ..and so.., ..the boy... One has then to specify how these are subsequently decomposed into a sequence of letter responses, as well as to show that such a process can be faster than finite state transcription.

Two hypotheses seem superficially plausible: one is that each familiar linguistic form, perceived as a unit pattern, is associated with a response unit which is an arpeggio of finger movements; the other is that each such linguistic unit has in its field of associations the string of letters that spells it, from which there is letter by letter transfer into response output.

The first hypothesis can be found in the work on morse transmission by Bryan and Harter (1899) and it surfaces again in the writings of Craik (1948). It assumes that for each response unit there is a fixed computing programme that, on instruction, co-ordinates a particular sequence of response elements. In extension there can be a hierarchy in which a programme at one level co-ordinates a sequence of programmes at a lower level (Miller, Galanter and Pribram, 1960). Even without such extension the hypothesis implies a large amount of learning in the training period, in which each linguistic form becomes associated with a finger pattern in keyboard language.

The hypothesis gains face validity from the observation that a touch typist usually performs not a sequence of independent finger movements but a flow of movements that modulate into one another. However it makes one prediction, which has not yet been tested but seems *a priori* unlikely. Note first that in the nature of the hypothesis, response programmes are associated only with their respective stimulus units (although they may be co-ordinated by higher programmes), hence there is no provision for transfer between them of instructions about common elements, except perhaps when one stimulus unit is embedded within another. Now suppose that two letters on the typewriter keyboard, say e and j, were interchanged, then the typist confronted with this change would, on the hypothesis, have to dismantle and reconstruct each programme in which one or other letter appeared. A test of this could be made most dramatic by retraining her with a restricted vocabulary and then comparing speeds with this and another vocabulary.

The other of the two hypotheses, discussed by Lashley (1961) and developed by MacNeilage (1964), would predict perfect transfer. We present a special version of the hypothesis, in which letter strings are translated into response sequences in two processing steps, the first imposing an ordinal labelling on the letters and the second mapping the letters into responses. The second step needs no comment beyond stressing that it confers no privileges for words and syllables at the response level, other than those made available by the first step.

The labelled letter strings yielded by the first step may be called the canonic form upon which the mapping rules can operate. We have in fact a very simple analogy to some grammars postulated by modern linguists to account for sentence production, in which syntactic markers provide a phrase structure for a morpheme string and phonological rules map this string into an utterance (Chomsky, 1957). Unlike here, however, the linguist is not concerned with the chronometric features of expression.

In transcription, reading and key pressing are asynchronous processes that are tied only to the same average rate. At any moment in time there is a queue of input elements awaiting translation, as indicated by the eye-finger span. The visual scan in reading is not a linear process but a sequence of saccadic jumps, each fixation covering a number of text elements that enter the queue together. Because these elements arrive simultaneously their original order may become lost, resulting in poor queue discipline, unless some ordinal labelling system is imposed to preserve it. A similar labelling notion was put forward by Yntema and Trask (1963) to account for some results in dichotic listening. Failures of queue discipline may be transmitted as transposition errors and even larger displacements of letters, e.g. ..at.. instead of ..ta.., or ..utalt.. instead of ..ualt...

A particular variant of transposition error is that in which the wrong letter in an (usually) adjacent pair gets repeated, e.g. ..aat.. instead of ..att... It suggests that double letters are stored as a single letter together with a repeat label which, like an ordinal label, may get displaced. Further analysis may yield still other forms of labelling device.

The labelling on letter strings also allows the mapping rules to operate in cyclic iteration, with consequent reduction of timing uncertainty, so that interresponse times in typing can be less than discrete simple reaction time measured under optimal conditions (cf. Shaffer, 1968).

Random letter strings require extensive improvised labelling while words and syllables that can be read as linguistic units already have labelled letter strings (their spellings) as associations and only the connections between them require improvising. If it takes longer to improvise labels than to retrieve existing ones then typing speed (and eye-finger span), which is limited by the rate of forming a canonic representation of text, will reflect this.

Context errors are probably more likely to occur with use of improvised labelling and so should be more frequent in degraded texts (see Table II). Their decline in L_0 is paralleled by our observation, confirmed by typists' comments, that for strings of uninterrupted consonants the letters have to be read into type one at a time, i.e. the queue reduces to zero. Context errors with prose and word texts tend to occur in long words but also occur in short words, usually of one syllable. Perhaps the typist attempts to encode the latter as part of larger linguistic fragments. We note that pairs of short words are occasionally typed in the wrong order, or the first or last letters of words may be displaced to corresponding positions in an adjacent word.

If a touch typist is asked where a certain letter is on the keyboard she may not be able to give the answer verbally, unless she goes through the finger dance of "typing" a word that includes it. It seems that finger movements are guided in keyboard space by reference to a movement sequence, and so any slowing of transcription may lead to a loss of reference. Thus there is an increase in Response and Omission errors in degraded texts. Such errors could also occur, of course, if the typist tries to go faster than she can accurately direct the trajectory of finger movement (many omission errors occurred because the key press was inadequate to bring the type in contact with the paper), but in this case they should decrease in degraded texts, as the typist goes slower, instead of increase (see Table II).

In short, our best answer so far to the question asked at the outset, what does the typist learn, is that she learns not the statistical structure of text (in any direct sense) but the use of a sequencing process that controls the orderly translation of linguistically encoded text into type.

We gratefully acknowledge the financial support of the Science Research Council for this research, which is part of a study on serial skills.

References

Bryan, W. L., and Harter, N. (1899). Studies on the telegraphic language. *Psychol. Rev.*, **6**, 345–75.

Butsch, R. L. C. (1932). Eye movements and the eye-hand span in typewriting. *J. educ. Psychol.*, **23**, 104–21.

Chomsky, N. (1957). *Syntactic Structures*. The Hague: Mouton.

Chomsky, N. (1963). Formal properties of grammars. Edited by Luce, R. D., Bush, R. R. and Galanter, E., *Handbook of Mathematical Psychology*, *Vol.* 2, New York: Wiley.

Conrad, R. (1964). Acoustic confusions in immediate memory. *Brit. J. Psychol.*, **55**, 75–84.

Craik, K. J. W. (1948). Man as an element in a control system. *Brit. J. Psychol.*, **38**, 142–8.

Genest, M. (1956). L'analyse temporelle du travail dactylographique. *Bull. Centre Étude Rech. Psychotech.*, **5**, 183–91.

Genest, M. (1957). Étude des fautes de frappe en dactylographie. *Bull. Centre Étud. Rech. Psychotech.*, **6**, 9–17.

Hershman, R. L., and Hillix, W. A. (1965). Data processing in typing. *Hum. Factors*, **7**, 483–92.

Lashley, K. S. (1961). The problem of serial order in behaviour. In Saporta, S. (Ed), *Psycholinguistics*, New York: Holt, Rinehart and Winston, 180–98.

MacNeilage, P. F. (1964). Typing errors as clues to serial ordering mechanisms in language behaviour. *Lang. Speech*, **7**, 144–59.

Miller, G. A., Galanter, E., and Pribram, K. H. (1960). *Plans and the Structure of Behaviour*. New York: Henry Holt.

Quastler, H., and Wulff, V. J. (1955). Human performance in information transmission, Part 2. *Rep. No.* 62, *Control Systems Lab., Univ. Illinois, Urbana.*

Schéffe, H. (1959). *The Analysis of Variance*. New York: Wiley.

Shaffer, L. H. (1968). Refractoriness in information processing. *Quart. J. exp. Psychol.*, **20**, 38–50.

Shannon, C. E. (1958). A mathematical theory of communication. *Bell System Tech. J.*, **27**, 379–423, 623–56.

Shannon, C. E. (1951). Prediction and entropy of printed English. *Bell System Tech. J.*, **30**, 50–64.

Sumby, W. H., and Pollack, I. (1964). Short-term processing of information. *HFORL Rep.*, *TR*-54-6.

Yntema, D. B., and Trask, F. P. (1963). Recall as a search process. *J. verb. Learn. verb. Behav.*, **2**, 65–74.

Manuscript received 4th May, 1968.

SHORTER ARTICLES AND NOTES

COGNITIVE CATEGORIES AND MEMORY SPAN: I. SHIFTING BETWEEN CATEGORIES

BY

A. F. SANDERS and J. J. F. SCHROOTS

From the Institute for Perception RVO-TNO, Soesterberg, The Netherlands

Broadbent and Gregory's hypothesis that shifting selection between cognitive classes is at least inefficient is tested by varying the number of transitions from consonants to digits and back in a mixed list memory span situation. No orderly decrease is found as a function of the number of transitions, which renders the attention hypothesis unlikely. An alternative position in terms of interference theory could explain the data qualitatively. Tests of further predictions of this theory need first an estimate of the degree of difference between cognitive classes.

INTRODUCTION

Broadbent and Gregory (1964) reported some years ago that it is harder to recall a list made up of alternate items of consonants and digits than it is to recall the same items arranged as two successive sublists, suggesting that going back and forth from one to another cognitive category is more difficult than remaining within the same class of items. An attention theory was proposed to explain this result: ". . . A class of items can be selected just as a sensory channel can and extra time is needed to shift this selection" (1964, p. 315). This idea is worked out in terms of Treisman's analysis on identification of incoming signals by matching them with some dictionary store of known words and symbols (Treisman 1960, 1964). The selection of a cognitive category, then, implies that the corresponding dictionary units have their thresholds lowered, relative to those of other classes. A shift of this response set to another cognitive category is supposed to be inevitably inefficient. So, the attention theory predicts an orderly decrease of recall as a function of the number of shifts from one to another category. Broadbent and Gregory only compared recall in the successive and alternate condition. It is the aim of the present paper to vary the number of transitions more systematically.

Method and procedure

Sixty lists of eight items each were constructed and recorded on sound tape at a presentation rate of two items per sec. The items used were the digits zero to nine or the consonants B, C, F, H, K, M, Q, R, T, X. No digit or consonant was repeated in a list. There were 10 lists for each of the following conditions: eight letters (L), eight digits (D), 1 Transition (LLLL–DDDD or DDDD–LLLL), 3 Transitions (LL–DD–LL–DD and vice versa), 5 Transitions (L–D–LL–DD–L–D and vice versa) and 7 Transitions (L–D–L–D–L–D–L–D and vice versa). Five lists of each of the four latter types started with L and five with D. The lists were presented in random order, and the subjects were given no information about the possible patterns.

Twelve officers of the Royal Netherlands Army, aged from 20 to 30 years, served as subjects. They were instructed to memorize the elements of the presented list and to recall them immediately afterwards by writing down on a response sheet. In the instructions special emphasis was given to ordered recall. In case of failures, subjects were advised to leave as many spaces as forgotten items of the sequence. Scoring was on the basis of correct items in correct positions.

Results

The mean number of correctly recalled items for each condition is presented in Table I. An analysis of variance showed a significant difference between conditions ($F = 11{\cdot}79$, $p < 0{\cdot}01$) and between subjects ($F = 13{\cdot}59$, $p < 0{\cdot}01$). A Newman–Keuls analysis indicated two significantly different clusters, one consisting of 8 L, 5 Tr and 7 Tr and the other of 8 D, 1 Tr and 3 Tr.

TABLE I

Mean Numbers of Correctly Recalled Items (out of 80)

		Number of transitions			
8 letters	*8 digits*	1	3	5	7
60·4	68·6	69·0	69·8	61·0	62·4

A comparison between recall of letters and digits in the 8 L, 8 D, 1 Tr and 3 Tr conditions showed that letters are better recalled at 1 Tr and 3 Tr than at 8 L, while digits are equally well recalled. In the same way, digits are less well recalled at 5 Tr and 7 Tr than at 8 D, while letters are not affected compared with 8 L. ($t = 2{\cdot}673$ $p < 0{\cdot}05$, and $t = 2{\cdot}795$; $p < 0{\cdot}05$ respectively.)

TABLE II

Mean Numbers of Correctly Recalled Letters and Digits in Mixed Lists

	Number of transitions							
	1		3		5		7	
	LLLL–DDDD	DDDD–LLLL	LL–DD	DD–LL	L–D	D–L	L–D	D–L
Letters	17·7	17·3	17·8	17·6	15·1	16·5	16·0	15·3
Digits	16·8	17·1	18·0	16·4	15·0	14·4	15·4	15·7

Discussion

The results agree with those of Broadbent and Gregory (1964) that recall of successive lists is superior to that of alternate lists. A simple attention switching theory is not supported however.

First, there is no monotonic decline as a function of the number of shifts. Secondly, recall of consonants is even better in moderately mixed lists (1 Tr, 3 Tr) than in comparable parts of pure lists of consonants, while the general decline of recall at 5 Tr and 7 Tr is solely due to less recall of digits. Both findings are difficult to explain, unless an *ad hoc* bias against digits is postulated. An alternative position could start from recent evidence on the importance of inter-item associations in ordered recall (Wickelgren 1965*a*, *b*). This assumes that the incoming signals activate not only corresponding units in the dictionary but also pre-existing associative bonds between successive units. When units have more specific and stronger traces and mutual associative connections, activation is likely to be more intensive during presentation of the list, such that intralist interference is better resisted. This refers clearly to the concept of differentiation in the interference theory of forgetting (e.g. Postman, 1963) and may be a major reason for the pronounced differences in memory span for various types of material (Mackworth, 1963; Sanders, 1968). So, digits appear to be more differentiated than the present set of consonants.

The principle of similarity may explain the recall in mixed lists. In the case of successive groups of items (1 Tr, 3 Tr), intralist interference is likely to be reduced, due to the presence of different types of activated units and associative bonds. The fact that digits are not better recalled in these conditions, may be an artefact of list length. When recall is almost perfect in a pure list of digits not much gain can be expected at 1 Tr, and 3 Tr.

However, when successive items belong to different categories, associative connections are likely to be weaker, since it can be assumed that bonds between items within classes are generally stronger or at least as strong as connections between items from different classes. This would imply that, when alternation prevails—as at 5 Tr and 7 Tr—weaker associations are activated most of the time. The decline at 5 Tr and 7 Tr is rather small, and recall still equals that of lists of consonants, indicating that consonants and digits have still quite strong connections, or that they do not constitute very different categories.

It is predicted that effects of mixed lists will be more pronounced when more different categories are involved. Reduction of interference should be larger in the case of 1 Tr, while the effect of weak interclass connections would be more pronounced in cases of alternation. These predictions are not provided by an attention switching theory. Experimental tests require quantitative estimates about differentiation and degree of difference of the categories involved. Differentiation might be derived from memory span. Difference is yet less feasible because there is no well defined criterion for deciding what constitutes a class of items. This general problem must be handled before predictions about shifting between categories can be considered.

REFERENCES

BROADBENT, D. E., and GREGORY, M. (1964). Stimulus set and response set: the alternation of attention. *Quart. J. exp. Psychol.*, **16**, 309–17.

MACKWORTH, J. F. (1963). The relation between the visual image and postperceptual immediate memory. *J. verb. Learn. verb. Behav.*, **2**, 75–85.

POSTMAN, L. (1963). Does interference theory predict too much forgetting? *J. verb. Learn. verb. Behav.*, **2**, 40–8.

SANDERS, A. F. (1968). Short term memory for spatial positions. *Ned. T. voor Psychol.*, **23**, 1–15.

TREISMAN, A. M. (1960). Contextual cues in selective listening. *Quart. J. exp. Psychol.*, **12**, 242–8.

TREISMAN, A. M. (1964). Selective attention in man. *Brit. Med. Bull.*, **20**, 12–6.

WICKELGREN, W. A. (1965*a*). Short term memory for repeated and non-repeated items. *Quart. J. exp. Psychol.*, **17**, 14–25.

WICKELGREN, W. A. (1965*b*). Acoustic similarity and retroactive interference in short term memory. *J. verb. Learn verb. Behav.*, **4**, 53–61.

Manuscript received 23rd April, 1968.

COGNITIVE CATEGORIES AND MEMORY SPAN: II. THE EFFECT OF TEMPORAL VS. CATEGORICAL RECALL

BY

A. F. SANDERS and J. J. F. SCHROOTS

From the Institute for Perception RVO-TNO, Soesterberg, The Netherlands.

SUMMARY

Subjects were presented with mixed lists of elements chosen from at least two formally distinct cognitive categories and were instructed to recall either temporally (i.e. in the actual order of arrival), or categorically (i.e. first all elements of one category followed by those of the other). Temporal recall is found to be superior when the categories are even and uneven digits; categorical recall is better when consonants and tones or consonants and spatial positions are mixed. Approximately equal performance is found with consonants and digits. It is suggested that efficiency of recall strategy depends on the relative strength of pre-existing associative bonds between elements from different cognitive categories. When these connections are weaker the probability increases that class membership rather than succession in time determines the assimilation of the list. It is found that with consonants and digits it does not matter whether instruction is given before or after presentation. This suggests that both recall modes can be activated at the same time. The implication of this result for work on dichotic listening is discussed.

INTRODUCTION

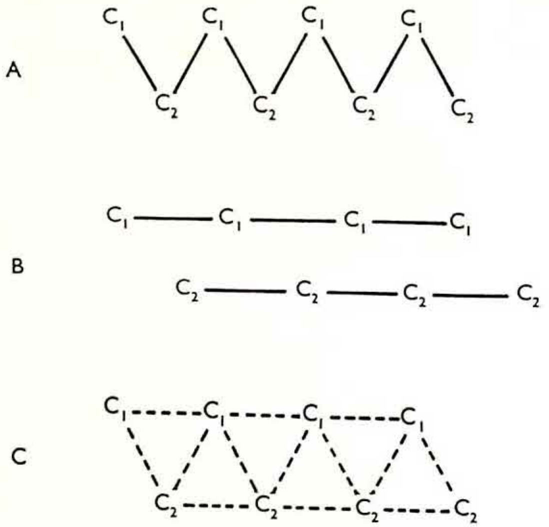
The distinction between various cognitive classes of signals, first studied by Külpe (1904), has appeared to be an important issue in human information processing and learning. The work of Von Restorff (1933) is classic in this respect and there is also much recent work especially in the field of attention and short term memory (Mackworth, 1963; Harrison, 1967; Yntema and Schulman, 1967).

Nevertheless it is still difficult to describe in more than a very loose sense what constitutes a cognitive category and what are its distinctive features from the point of view of information processing. Letters and digits, for instance, have been traditionally considered as distinct classes, but one may think of more elaborate subdivisions, like vowels and consonants or odd and even digits, but there is no test to estimate the behavioural importance of such classifications or to measure the relative distances between categories.

Several approaches are possible to this problem. One is based on the evidence that a conditioned response to a word is easier to generalize to other words with similar meaning than to words of related sound, suggesting an effect of categorization (Luria and Vinogradova, 1959). Another possibility is the investigation of clustering of responses of the same category in free recall of lists of items, made up of randomly arranged elements of different categories (Bousfield, 1953; Tulving, 1962).

The theoretical background of both approaches is in associationist terms, assuming that incoming material activates pre-existing traces and associative connections in a dictionary store of known symbols. A further assumption is that associative bonds within categories are generally stronger than between classes, especially when the categories differ widely. This is also the idea behind the approach, outlined in the following experiments, which makes use of a memory span context with lists containing items of two cognitive classes. It is characteristic of memory span studies that both order and event information are important so that, apart from memorizing items, short term retention of associations between successively presented items is essential (see Fig. 1A). When more remote cognitive categories are mixed in one list the chance would increase however that not succession in time but membership of the same class determines which association is activated (see Fig. 1B). It follows then that when recall in temporal order is required, performance will be superior if there is less difference between categories, whereas recall

FIGURE 1



Possible associative connections in a list made up of alternate items of two classes.

in categorical order will be better when the differences are larger. An interesting point may be when the difference of between and within class associations approximately matches the effect of succession in time, which implies that temporal and categorical recall are performed equally well (see Fig. 1C). Such has been found with dichotic presentation of digits and nonsense syllables or consonants, although the temporal course is distorted in that situation (Broadbent and Gregory, 1964; Yntema and Trask, 1963; Gray and Wedderburn, 1960). It is interesting therefore to know about temporal and categorical recall of digits and consonants in the monaural case. Other tests concern consonants and pure tones, and consonants and visual spatial positions. Both tones and positions belong to non-verbal categories and may have weak associative connections with consonants. Finally, odd and even digits are used as formally distinct cognitive classes, under the assumption that they are hardly different from the point of view of handling information.

EXPERIMENT I

Method and Procedure

Lists of eight items were prepared for each combination of categories. They were recorded on sound tape, with the exception of the letters-spatial positions condition, where a visual presentation technique was employed. Three types of lists were constructed, i.e. three transitions between categories (3 Tr: AA-BB-AA-BB), five transitions (5 Tr: A-B-AA-BB-A-B) and seven transitions (7 Tr: A-B-A-B-A-B-A-B). In all conditions except the odd vs. even digits, there were 20 lists of each Tr type, making a total of 60 lists for each group of subjects. In the odd vs. even digits condition only 3 Tr and 7 Tr lists were presented, with 12 lists of both types, making a total of 24 lists, for that group. The order was always randomized. All combinations of categories were taken by two groups of subjects (with the exception again of the odd vs. even digits). One of these groups was informed before a presentation how many transitions would be in the next list (Knowledge of Tr). During the instruction they were extensively informed about the possible structures, which remained on continuous display on a blackboard during the experiment. No knowledge of Tr was given to the other group. Half of the lists of each Tr type were temporally recalled, i.e. in the actual order of arrival of the items, and the other half was recalled categorically, i.e. first all items from the category with which the list started, followed by the material from the remaining category. Within categories the order of arrival had to be maintained. The instruction before presentation of a list specified which type of recall was required.

In all conditions where auditory presentation was used, subjects were tested as a group and wrote the recalled items on a response sheet. The reproduction interval amounted to 20 sec. Final details of instructions concerned emphasis on immediate recall after presentation of a list and information about presentation rate (2 items/sec). Subjects were also told that scoring would be on the basis of correct item in correct position. In case of partial forgetting they were advised to leave open those spaces where the lost items were located, or to guess. Other details of method and procedure apply only to specific conditions.

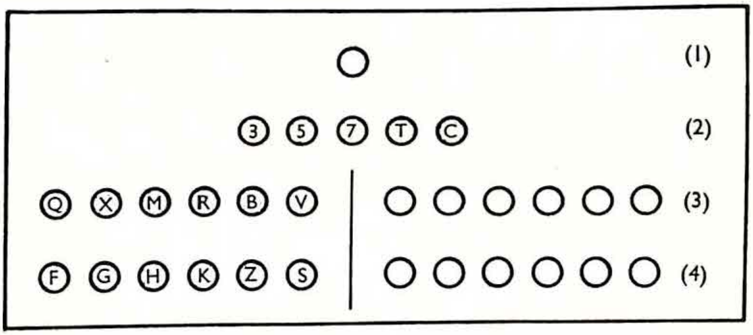
(1) *Letters–digits.* All digits from one to nine were used. The following subset of consonants was used: B, C, F, H, K, M, Q, R, T, X. They were selected to minimize auditory confusion (cf. Conrad and Hull, 1964) and were on permanent display during the experimental sessions. Two groups of 10 freshmen from the University of Utrecht were tested.

(2) *Letters–tones.* The same set of consonants was used. The tones consisted of a clearly audible high and low pitched tone, differing one octave (262 Hz and 524 Hz). Only two tones were employed to avoid absolute judgement problems (e.g. Pollack, 1953). Subjects indicated the high pitch tone with an upward arrow (↑) and the low pitch tone with a downward arrow (↓). Two groups of 12 freshmen served as subjects.

(3) *Odd and even digits.* 0, 2, 4, 6, 8 were used as even digits, and 1, 3, 5, 7, and 9 as odd digits. Special emphasis was given in the instruction about the use of the digit 0 as an even number. Nine academic laboratory employees served as subjects and no knowledge of Tr was given.

(4) *Letters–spatial positions.* Letters and spatial positions were visually displayed in a set up, which consisted of 30 lights on a panel, as schematically presented in Fig. 2.

FIGURE 2



Schematic drawing of the visual display.

The whole programme of lists and reproduction intervals was prepared on punched tape and fed into an optical reader. The interval and event codes were then translated in an electronic timing device, activating the appropriate signals and intervals (Van Doorne and Sanders, 1968). Subjects were seated in front of the display, covering a visual angle of 60°. Four sec. before presentation of a list, the top row light (1) flashed for 1 sec. After an interval of 1 sec. followed the instruction "temporal" or "categorical" by means of a 1 sec. flash of the lights T or C (row 2). In the case of information about the number of transitions in the list, the lights indicating 3, 5 or 7 flashed simultaneously with the instruction light. After another 1 sec. interval the sequence of items was given by flashing lights from the rows 3 and 4. The consonants B, F, G, H, K, M, Q, R, S, V, X, Z were mounted on the lights in the left part of the display, while in the right part only simple light flashes were given (see Fig. 2). Subjects were instructed to memorize the consonants and the spatial position of the simple light flashes. Each flash lasted 500 msec. and was followed immediately by the next one. Symbols were only visible during a flash. Subjects recalled the consonants by speaking aloud and the spatial positions by pressing the appropriate response keys. To obtain optimal S–R compatibility the lights had been mounted on top of response keys, so that recall of spatial positions consisted of touching the lights in order of presentation (Fitts and Biederman, 1965). Recall was registered by the experimenter and the key reactions were recorded on a digital printer which indicated which light had been touched.

This set up is very similar to the one that was used in some experiments on STM for spatial positions (Sanders, 1968). It was found there that a presentation rate of two signals per sec. presented no difficulties in keeping track of the signals, at least not for a list of eight items. Before the experimental session subjects were given a trial in which they had to reproduce each item upon arrival and this proved quite possible within the limits of the length of a list.

All subjects had six training trials, which were repeatedly presented and recalled, until subjects showed that they had adapted to the situation. In all other conditions two training trials were given for each variable. Seven freshmen served in the knowledge of Tr group, while eight laboratory employees were used in the condition where no knowledge of Tr was given.

Results

(1) Mean performance per list was calculated as a function of the main variables of the experiment: combination of categories, number of transitions (Tr), temporal (T) and categorical (C) recall, and knowledge of Tr. The data are presented in Table I.

TABLE I

MEAN NUMBERS OF CORRECTLY RECALLED ITEMS PER LIST

	Knowledge of Tr						*No knowledge of Tr*					
	3 *Transitions*		5 *Transitions*		7 *Transitions*		3 *Transitions*		5 *Transitions*		7 *Transitions*	
Combination of categories	T	C	T	C	T	C	T	C	T	C	T	C
Digits: even–uneven	—	—	—	—	—	—	5·50	5·28	—	—	5·85	4·58
Letters–digits	7·26	7·07	5·80	5·76	6·40	5·69	6·11	6·06	5·50	4·88	5·33	5·01
Letters–tones	6·79	7·45	5·81	6·20	5·96	6·67	7·02	7·54	5·55	6·52	5·48	6·91
Letters–positions	5·46	6·81	3·94	5·54	3·93	5·99	4·05	5·60	3·01	5·01	3·21	5·32

T = Temporal. C = Categorical.

An analysis of variance, followed by a Newman–Keuls analysis for post hoc comparison of individual means, was carried out for each experimental group. The results are summarized in Table II.

TABLE II

SURVEY OF SIGNIFICANT F-VALUES

	Knowledge of Tr			*No knowledge of Tr*			
Factors	*Letters–digits*	*Letters–tones*	*Letters–positions*	*Letters–digits*	*Letters–tones*	*Letters–positions*	*Digits: even–uneven*
Number of Tr (A)	47·41 < 0·001	23·52 < 0·001	19·59 < 0·001	18·17 < 0·001	36·05 < 0·001	7·85 < 0·01	N.S
Temporal/Categorical (B)	N.S	18·22 < 0·001	72·79 < 0·001	5·76 < 0·05	12·03 < 0·01	53·49 < 0·001	20·74 < 0·001
Subjects (C)	11·56 < 0·01	5·73 < 0·01	14·06 < 0·001	18·92 < 0·001	9·72 < 0·01	29·98 < 0·001	17·09 < 0·01
A × B	3·80 < 0·05	N.S	N.S	N.S	4·00 < 0·05	N.S	6·01 < 0·05
A × C	N.S	N.S	N.S	N.S	4·68 < 0·01	2·44 < 0·05	N.S
B × C	N.S	N.S	N.S	N.S	N.S	N.S	N.S

It appears that recall at 3 Tr is always better than at 5 or 7 Tr, except at the odd vs. even digits condition. Furthermore, there appear minor differences between temporal and categorical recall in the letters–digits groups. Temporal recall is much better with odd and even digits, and worse with tones and still worse with spatial positions.

(2) A clear idea of the differences between the experimental groups can be obtained by taking the logarithms of the ratio of temporal and categorical recall for each subject. A zero score, then, means equal performance with both instructions. The data are shown in Table III.

TABLE III

MEAN ($\log_{10}$ RECALL CATEGORICAL—$\log_{10}$ TEMPORAL) FOR THE VARIOUS CONDITIONS

	Digits odd–even	*Letters–digits*	*Letters–tones*	*Letters–positions*
Knowledge of Tr	—	−0·05	+0·07	+0·22
No knowledge of Tr	−0·11	−0·03	+0·12	+0·27

The value 0 means equal categorical and temporal performance. All data are from lists with 7 Tr.

An analysis of variance on these data revealed highly significant differences between conditions ($F = 27{\cdot}93$, $p < 0{\cdot}01$). *Post hoc* comparisons of individual means did not show evidence for a significant effect of knowledge of the number of transitions. The differences between the various combinations of cognitive categories were all significant.

(3) Separate analyses were carried out to decide whether differences between temporal and categorical recall might be specifically due to one of the cognitive categories. No systematic evidence was found however.

Discussion

The results confirm the idea that temporal order in memory span situations is not always a dominant factor in the efficiency of recall. In fact, the efficiency of temporal versus categorical recall in mixed lists depends strongly on which categories are used. The relative efficiency does agree well with intuitive ideas about differences between cognitive categories. Temporal recall is best at the odd and even digits and, moreover, this is the only condition in which no difference in recall appeared between 3 Tr and 7 Tr. It can be inferred from previous work that differences between 3 Tr and 7 Tr occur when the items come from cognitive classes which are functionally different (Broadbent and Gregory, 1964; Sanders and Schroots, 1968). The conclusion may be therefore that the distinction between odd and even digits is not relevant from the point of view of handling information but has only a formal basis.

The experimental groups with knowledge of Tr were introduced to control for a possible bias in favour of categorical recall. Subjects then know at which places to reproduce items of a given category in the case of categorical recall and this knowledge may positively affect recall. This bias is compensated for with the "knowledge of Tr" groups. The results show indeed a non-significant although systematic tendency towards relatively better temporal recall with the latter groups, although the effect is slight and does not seem to affect the main argument.

It is interesting to note that, as with dichotic lists, temporal and categorical recall are performed about equally well in the monaural lists of digits and consonants used in these experiments. This raises the question whether either temporal or categorical associations are activated—on the basis of instruction for instance—or whether both can be aroused at the same time. The selective strategy at the input stage is decisive, given the first hypothesis and it would lead to an explanation in terms of a response set, as was proposed by Broadbent and Gregory (1964) to explain the data on dichotic listening. In the second case, it would be possible to follow either the temporal or the categorical line when instruction is given after presentation of a list. This would rather lead to the search theory of Yntema and Trask (1963), provided that the search is guided by the associative connections and the instruction after presentation. Categorical recall, therefore, is not necessarily a matter of input selection, but may be bound to retrieval, in which case the tricky problem of semantic selection criteria of the selective input mechanism is avoided. This question is investigated in the following experiment, using monaural mixed lists of consonants and digits.

EXPERIMENT 2

Method and Procedure

Sixty lists of eight items each were constructed and recorded on sound tape at a presentation rate of two items per sec. The same consonants and digits, the same compositions (3 Tr, 5 Tr, 7 Tr) were used as in the first experiment, and the lists were temporally or categorically recalled. In half of the trials the instruction as to recall mode was given before presentation of the list and in the other half instruction was given after presentation of the list. The order of presentation was randomized. Since instruction after presentation is likely to interfere with retention of the list, another binary choice task was introduced. Subjects had two response sheets, marked L and R respectively, and were instructed on which sheet to reproduce a list. When recall mode instruction was given afterwards, the instruction concerning L and R was given before presentation. In the same way L or R was instructed after presentation if recall mode was mentioned before a list was given.

Nine laboratory employees, who had not taken part in an earlier study of this project, served as subjects. The usual instructions and two training trials were given and these were repeated and discussed until all subjects felt sure that they were familiar with the task. Further details of instructions, procedure and scoring were the same as in the previous experiment.

Results

The results are summarized in Table IV. They were subjected to an analysis of variance. There was no significant evidence for a difference between temporal and categorical recall nor for instruction of recall mode before or after presentation. Number of transitions was highly significant, as in the previous experiments ($F = 26{\cdot}09$, $p < 0{\cdot}01$).

TABLE IV

MEAN NUMBER OF CORRECTLY RECALLED ITEMS PER LIST

		Temporal			*Categorical*		
		Number of transitions			*Number of transitions*		
		3 Tr	5 Tr	7 Tr	3 Tr	5 Tr	7 Tr
Instruction concerning recall mode	Before presentation	4·26	3·08	3·56	4·60	3·58	3·74
	After presentation	5·08	3·72	3·64	4·44	3·80	3·78

Discussion

The results show that the interfering effect of instruction about recall order after presentation does not exceed the effect of choosing between writing the response on the left or right response sheet. It is probable, therefore, that recall order can be efficiently chosen after presentation of the list, which suggests that both the temporal and categorical line can be activated during presentation of a list. This explanation can be also tentatively offered to explain the results on dichotic listening with mixed lists, although a comparable experiment must be carried out within a dichotic experimental setting.

Taking the results of the experiments together, it can be said that the measurement of temporal and categorical recall can be at least one approach to the problem of cognitive categories. They do not support a simple theory of switching attention betweeen categories (e.g. Broadbent and Gregory, 1964), since it is hard to explain in these terms why it would be less difficult to switch from consonants to digits than from consonants to spatial positions. The data rather point to pronounced differences in strength of associative connections between various cognitive categories. There are of course a number of variables, like visual vs. auditory presentation, written and spoken recall and rate of presentation, which are likely to exert an effect on the activation and retrieval of the material and the associative bonds, and which are not effectively controlled in the present study. The determination of the relative weights of these effects is not only important for the proposed measure of cognitive categories but also for the knowledge about information processing in short term memory tasks.

Finally the results of the present measure would need to be found to be consistent with those of other earlier mentioned approaches, before the concept of cognitive category can change from an operational to a theoretical construct.

REFERENCES

BOUSFIELD, W. A. (1953). The occurrence of clustering in the recall of randomly arranged associates. *J. Gen. Psychol.*, **49**, 229–40.

BROADBENT, D. E., and GREGORY, M. (1964). Stimulus set and response set: The alternation of attention. *Quart. J. exp. Psychol.*, **16**, 309–17.

CONRAD, R., and HULL, A. J. (1964). Information, acoustic confusion and memory span. *Brit. J. Psychol.*, **55**, 429–32.

DOORNE, H. VAN, and SANDERS, A. F. (1968). A programmed stimulus and response processor. *J. Behav. Res. Meth. Instr.*, in press.

FITTS, P. M. and BIEDERMAN, I. (1965). S–R compatibility and information reduction. *J. exp. Psychol.*, **69**, 408–12.

GRAY, J. A., and WEDDERBURN, A. A. J. (1960). Grouping strategies with simultaneous stimuli. *Quart. J. exp. Psychol.*, **12**, 180–5.

HARRISON, G. J. (1967). Some additive results in short term memory. In: SANDERS, A. F. (Ed.), Attention and performance. *Acta Psychol.*, **27**, 306–316.

KÜLPE, O. (1904). Versuche über Abstraktion. *Ber. 1. Kongress Exp. Psychol. Berlin*, 56–8.

LURIA, A. R., and VINOGRADOWA, D. S. (1959). An objective investigation of the dynamics of semantic systems. *Brit. J. Psychol.*, **50**, 89–105.

MACKWORTH, J. F. (1963). The relation between the visual image and post-perceptual immediate memory. *J. verb. Learn. verb. Behav.*, **2**, 75–85.

POLLACK, I. (1953). The information of elementary and auditory displays. *J. acoust. soc. Amer.*, **25**, 765–9.

SANDERS, A. F. (1968). Short term memory for spatial positions. *Ned. T. Psychol.*, **23**, 1–15.

SANDERS, A. F., and SCHROOTS, J. J. F. (1968). Cognitive categories and memory span. I. Shifting between categories. *Quart. J. exp. Psychol.*, **20**, 370–2.

TULVING, E. (1962). Subjective organisation in free recall of "unrelated" words. *Psychol. Rev.*, **69**, 344–54.

VON RESTORFF, H. (1933). Uber die Wirkung von Bereichsbildungen im Spurenfeld. *Psychol. Forsch.*, **18**, 299–351.

YNTEMA, D. B., and TRASK, F. P. (1963). Recall as a search process. *J. verb. Learn. verb. Behav.*, **2**, 65–74.

YNTEMA, D. B., and SCHULMAN, G. M. (1967). Response selection in keeping track of several things at once. In: SANDERS, A. F., (Ed.), Attention and performance. *Acta Psychol.*, **27**, 325–35.

Manuscript received 23rd April, 1968.

DISTRACTION AND TIME ESTIMATION

BY

GUY VON STURMER, TONG WONG and MAX COLTHEART

From the Department of Psychology, Monash University

It is argued that events which occur during an interval of time which is being judged may be classified in terms of their effects on the alertness of the subject, and in terms of the degree to which they distract him from the task of detecting and processing time-relevant cues. A distracting task, defined by the number of arithmetical operations a subject was required to perform, was presented while reproductions of an interval were being made. The data support the prediction that the higher the level of distraction, the less time a subject will judge to have elapsed during an objective period.

INTRODUCTION

In a discussion of the contradictory literature on filled versus unfilled time Fraisse (1963) suggests that the more "attention" is paid to the task of judging the duration of an interval, the more time will be judged to have elapsed. The concept of attention has two main aspects. It involves the notions of alertness, or vigilance, and selectivity. In a recent paper (von Sturmer, 1966) the data presented supported the view that the more alert a subject is in his task of judging durations, the more time he will judge to have elapsed during a given interval.

This paper is concerned with the distracting effect of events which occur during an interval which is being judged. It is clear from work summarized by Berlyne (1960) that stimulus properties which increase alertness may also simultaneously distract the subject from attending to other stimulus properties. In time estimation experiments it may be easier to disentangle these effects if a dominantly distracting situation is regarded as one where most or all of the subject's capacity to respond is occupied with something other than judging duration.

The distracting power of a task can be independently measured in terms of the time a subject takes to perform it, i.e. time during which the subject is assumed to be doing something other than detecting and processing time-relevant cues. Alternatively, distracting power can be measured in terms of the number of operations the task demands to be performed.

The present theoretical view of the filled versus unfilled time problem may be most simply described as involving two factors—alertness and response competition. Increases in alertness result in intervals being judged longer. Increases in distraction result in intervals being judged shorter. Implicit in this are the assumptions that (*a*) the more cues for lapse of time (suprathreshold changes in the perceptual system) are detected, the more time is perceived to elapse, and (*b*) the more time there is available to make inferences of time lapse from these cues, the more time will be perceived to elapse.

In the experiments reported here distraction was varied by having the subject perform arithmetical tasks, involving the addition of various numbers of digits, while making reproductions of a 10 sec. standard interval. It was expected that the more digits the subject had to add while making his judgement, the longer the judgement would be—indicating that less time was being perceived to elapse, per objective unit of time, during the reproduction.

The degree to which this task was distracting was also varied for different groups of subjects by instructing them variously that the time estimation task was the most important, the arithmetical task was the most important, or the two tasks were of equal importance. It was expected that the distracting effect would be greatest with the group which was told that the arithmetical task was the most important, and least in the group which was told that the time estimation task was the most important.

EXPERIMENT I

Procedure

Training. The subject was told that he was to be trained to make accurate judgements of 10 sec. He was seated at a tachistoscope and the field was switched by a timer to expose an *X* on a white ground for 10 sec. The subject was told that the duration of the exposure of the X was the standard interval. He was then instructed to press the manual control switch when the experimenter said "now" (10 sec. after offset of the standard) and to keep it depressed for 10 sec. without counting. While the switch was depressed by the subject the card with X on a white ground was again exposed, this time for the duration of his judgement. After the judgement had been recorded with an electronic timer the subject was told what his estimate was and the procedure was repeated with a 15-sec. intertrial interval. Training ended when the subject had made 30 judgements with knowledge of results, or, in the case of very accurate subjects, earlier than this if the subject was able to make five consecutive judgements without knowledge of results and within 10 per cent. of the standard. The purpose of the training procedure was to acquaint the subject with the apparatus and to gain a greater degree of uniformity of initial estimates in the test situation.

Practice. Five practice trials were then given. These differed from the test trials which followed only in that the subject was allowed to ask questions about procedure and the data were not used in the statistical analysis. No knowledge of results was given.

On practice and test trials the standard was presented as before—an X at the centre of a white ground exposed for 10 sec. Similarly, the judgement was started 10 sec. after offset of the standard. The intertrial interval, however, was 35 sec.

During judgements in practice and test sessions depression of the manual control switch by the subject exposed a card bearing a number of digits. The subject was instructed to add these while making his judgement.

The digits were printed on white cards which were slipped into the tachistoscope between offset of the standard and onset of the judgement. The numbers of digits on the cards corresponded to each of five levels of distraction. Each card bore 0, 4, 8, 16 or 32 digits taken from a random number table. The subject had to add aloud, giving the cumulative sum each time he read a digit. When the subject released the control switch at the end of a time judgement the stimulus card vanished. Each level of distraction was presented once during practice.

Subjects were told that the stimulus card must not be retained in view after the estimated 10 sec. had elapsed. If some digits remained to be added the card was to be released at 10 sec. and the total of the remaining digits could be guessed.

Testing. After the practice run the subject was given instructions designed to vary the effectiveness of the distracting conditions. One group was told that, although time estimation and addition were required to be done simultaneously, the accuracy of the time judgements was the most important feature of the experiment. A second group was told that both tasks were of equal importance, and the third group was told that the accuracy of the arithmetical task was of prime importance.

Each subject made five judgements at each of the five levels of distraction in the test situation. On each occasion, at each level, a different random order of digits was presented on the stimulus card for each subject. No subject saw the same stimulus card twice except for the standard card and the card for the zero level which simply bore an X as a focal point. The order in which the trials were given was randomized for each subject.

Subjects

Thirty-six university students were used as subjects. Approximately equal numbers of men and women were assigned to each group randomly. None of the subjects knew the purpose of the experiment.

Results

From Table I it can be seen that the reproductions of 10 sec. increased in duration as the number of digits to be added was increased.

Each subject gave five scores at each level of distraction. When averaged for each subject the data were markedly skewed and the variances were heterogeneous. Accordingly non-parametric tests were used to examine the effects of distraction levels.

A non-parametric test of trend for correlated groups (Ferguson, 1965, ch. 4) was used to test the hypothesis that duration of reproduction would be a positive monotonic function of level of distraction. This hypothesis was confirmed for each of the three instructional groups (z-values of $+ 6{\cdot}1$, $+ 7{\cdot}1$ and $+ 6{\cdot}4$ being obtained).

An analogous non-parametric test of trend for independent groups (Ferguson, 1965, ch. 7) was used to test the hypothesis that the arithmetic-emphasis group would yield the greatest durations of reproductions, and the time-emphasis group the least, with the

TABLE I

MEANS AND STANDARD DEVIATIONS (IN SEC.) AT EACH LEVEL OF DISTRACTION UNDER THREE INSTRUCTIONAL CONDITIONS

Instructions		*No. of digits to be added*				
		0	4	8	16	32
A: time emphasis	Mean	10·7	10·0	11·4	15·0	19·0
	S.D.	2·2	2·4	2·6	3·6	7·2
B: neutral	Mean	10·4	11·2	12·7	16·4	20·8
	S.D.	1·3	2·4	2·6	4·4	7·4
C: arithmetic emphasis	Mean	11·5	11·6	12·8	18·7	25·5
	S.D.	1·7	2·5	1·9	4·0	8·6

Twelve different subjects were used in each instruction condition.

neutral group intermediate. This analysis was performed separately upon the data obtained with each level of distraction. None of these five anaylses yielded any evidence of a relationship between instructional condition and duration of reproduction ($z < 1{\cdot}0$ in all cases).

Discussion of Experiment 1

The trend analysis supported the prediction that the more digits a subject was required to add while reproducing 10 sec., the longer his judgement would be. Although the means were in the predicted directions, the instructions, which were intended to differentially emphasize the time and the arithmetic task, did not produce a statistically significant effect. This is possibly only because the subjects were not as credulous as was originally hoped. Where subjects were told to pay more attention to one task than to the other it was still obvious that performance on both tasks was being measured. It was not altogether surprising, therefore, that this secondary aspect of the study yielded insignificant data.

EXPERIMENT 2

Although the results of the first experiment tend to support the distraction hypothesis, it could be argued that the lengthening of reproductions which occurred with increases in the number of digits to be added came about because of a tendency for subjects to keep the stimulus card in view after 10 subjective sec. had elapsed in order to do better on the arithmetical task. This may have occurred in spite of the fact that subjects were told to guess the total of the digits remaining to be added after time judgements had been completed. A second experiment was designed, therefore, in which the digits remained in view *after* the subject had signalled the end of his time judgement.

Procedure

In this study the experimenter simultaneously switched on a recording timer and the tachistoscope displaying the addition task. The subject was told to begin adding (reading cumulative totals aloud) the moment the display was switched on. His reproduction of 10 sec. was to begin at the same time. At the end of the time judgement the subject tapped a Morse key which stopped the timer. The digits remained in view, however, and the subject continued to add them until the arithmetical task was completed. When both tasks were completed the experimenter switched the tachistoscope off. Under these conditions the subject was forced to carry out both tasks simultaneously until his time judgement had elapsed, then continued adding if necessary. He could not expect to increase his accuracy in arithmetic by prolonging his judgement of 10 sec.

Twelve subjects were used in a procedure which was otherwise essentially the same as that for the neutral group in the first study. To reduce the time it took to run each subject, however, a criterion of three successive trials without knowledge of results and within 10 per cent. of the standard was used during training. The inter-trial interval during testing was 30 sec. and the interval between the offset of the standard and the onset of the judgement was timed with a stopwatch at approximately 2 sec. The number of digits displayed on each trial was either 4, 8, 12, or 16.

Results

From Table II it can be seen that the effect of the distracting task was similar to that obtained in the first experiment.

TABLE II

MEAN JUDGEMENTS (IN SEC.) AT EACH LEVEL OF DISTRACTION IN EXPERIMENT 2

	No. of digits to be added			
	4	8	12	16
Mean reproductions	10·0	11·3	12·8	14·4
Standard deviations	2·1	2·4	3·5	4·2

Hartley's test for heterogeneity (Winer, 1962) indicated that the assumption of homogeneity of variance was satisfied here. A trend analysis (Edwards, 1967, ch. 15) showed that the duration of judgements increased linearly with level of distraction ($F = 54{\cdot}20$, $p < 0{\cdot}001$) and that the residual (non-linear) trend was not significant ($F < 1{\cdot}0$).

DISCUSSION

These two experiments support the distraction hypothesis which is outlined in the Introduction. The second experiment tends to rule out the most obvious alternative explanation which could be given for the results of the first study.

In previous studies of judgements of filled versus unfilled time, where the variables can be classified according to the definition of distraction given in the Introduction, the data agree with the present theoretical views (e.g. Harton, 1938*a*, *b*; Bakan, 1955; DeWolfe and Duncan, 1959; Banks and Cappon, 1962; Warm, Greenberg and Dube, 1964). Other confirmatory evidence is summarized under various headings in Wallace and Rabin (1960) and Fraisse (1963).

In these studies the range of explanatory concepts used is almost as wide as the variation in experimental variables employed as "fillings." None of the various concepts put forward previously has both a wide range of applicability and lack of ambiguity in definition. The present system for classifying events occurring during a judged time interval seems to have the two advantages of operational specificity and generality of application. It is also shown here to have some predictive utility.

There appears to be no reason why the present approach could not be profitably applied in studies of judgements of other features—such as extensity and numerosity. A greater degree of precision may be attained, however, if distraction is quantified in information theory terms.

REFERENCES

BAKAN, P. (1955). Effect of set and work speed on time estimation. *Percept. mot. Skills*, **5**, 147–8.

BANKS, R. and CAPPON, D. (1962). Effect of reduced sensory input on time perception, *Percept. mot. Skills*, **14**, 74.

BERLYNE, D. E. (1960). *Conflict, Arousal and Curiosity*. New York: McGraw-Hill.

DEWOLFE, R. K. S., and DUNCAN, C.P. (1959). Time estimation as a function of level of behaviour of successive tasks. *J. exp. Psychol.*, **58**, 153–8.

EDWARDS, A. L. (1967). *Statistical Methods*. New York: Holt, Rinehart and Winston (2nd ed.).

FERGUSON, G. A. (1965). *Nonparametric Trend Analysis*. Montreal: McGill University Press.

FRAISSE, P. (1963). *The Psychology of Time*. New York: Harper and Row.

HARTON, J. J. (1938*a*). The influence of the difficulty of activity on the estimation of time, I. *J. exp. Psychol.*, **23**, 270–87.

HARTON, J. J. (1938*b*). The influence of the difficulty of activity on the estimation of time, II. *J. exp. Psychol.*, **23**, 426–38.

WALLACE, M., and RABIN, A. I. (1960). Temporal experience. *Psychol. Bull.*, **57**, 213–36.

WARM, J. S., GREENBERG, L. F., and DUBE, C. S. (1964). Stimulus and motivational determinants in temporal perception. *J. Psychol.*, **58**, 243–8.

VON STURMER, G. (1966). Stimulus variation and sequential judgements of duration. *Quart. J. exp. Psychol.*, **18**, 354–7.

WINER, B. J. (1962). *Statistical Principles in Experimental Design.* New York: McGraw-Hill.

Manuscript received 17th June, 1968.

FAMILIARITY AND WORD RECOGNITION

BY

R. F. GARTON and L. R. ALLEN

From the Department of Psychology, University of Melbourne

Word recognition was examined for physics and non-physics students after a single visual presentation of a series of sixty words. Physics words likely to be familiar only to physics subjects were used to vary the prior familiarity of the two groups with the inspection material and the ratio of physics to common words (salience) was varied for each of the three inspection series. Common words, which were expected to have greater salience than physics words in one of the inspection series, were not found to be recognized better than the control pairings of physics words by either group of subjects. The absence of differences between physics and non-physics students in the number of inspection list words (positive instances) correctly identified supports the view that prior familiarity is not a critical variable when all inspection items are presented in a recognition test. Physics students made fewer errors than non-physics students in identifying both physics and common words in the the recognition test which were not in an inspection series (negative instances). It is suggested that the correct identification of negative instances involves a scanning of a subject's "internal store" of previously experienced inspection items and that this is more efficient when the meaning of all the items in this "store" is known.

INTRODUCTION

The role of familiarity in the recall of complex material has interested psychologists since the classic work of Bartlett (1932). He stressed that recall was directly influenced by the way in which a person organized and interpreted perceived material. The importance he attributes to existing cognitive structures ("schemata") is clear from his statement that "a particular response is possible only because it is related to other similar responses which have been serially organized, yet which operate, not simply as individual members coming one after another, but as a unitary mass" (p. 201). This cognitive function of ordering incoming information has been elaborated by Bruner (1957) in terms of categorizing stimulus input. The extent to which a person is equipped through past experience with appropriate modes of coding information determines his ability to organize his environment in a variety of cognitive tasks.

It may also be argued, however, that the *ratio* of familiar to unfamiliar material is of importance. If for example the task is the recognition of words presented visually, the proportions of familiar and unfamiliar words may be varied. In a series such as "religion, soldier, difference, dynatron, government," the word "dynatron" may be subsequently recognized with ease even if it is not part of the subject's everyday vocabulary. Such an observation could be regarded as an example of the von Restorff (1933) effect.

A modern analogy of this effect would be one borrowed from Signal Detection Theory (Swets, 1964). One may view "dynatron" in the above example as a more discriminable "signal" amid a series of signals of common words. Similarly, in a series such as "solenoid, tetrode, hysteresis, religion, dynatron" perhaps "religion" would have a higher threshold and be recognized more easily. Though a common word, its presence in this list is unusual. Consequently, this von Restorff effect would cover the cases where familiar and unfamiliar words may stand out depending upon which of the two classes was in the minority.

The viewing of familiar and unfamiliar words within a context of other words may be regarded as a problem in detection. A subject knows that in a recognition test he may be confronted by a signal (i.e. a previously presented item) or by noise (i.e. extraneous words), and he is to report which he believes it to be (see Luce, 1959; Green and Swets, 1966). Alteration of the ratio of familiar to unfamiliar items in an inspection list may induce a variation of recognition strategy.

The exploratory investigation reported here examines the relative effects of familiarity and salience when these attributes are varied systematically in a word-recognition task.

METHOD

Subjects. The subjects were 150 university students, 75 having completed at least one year of a physics course and 75 having little, if any, science training at secondary school.

Materials. Three inspection lists were prepared consisting of physics and non-physics words.

The physics words (P) were chosen in order that they would be very familiar to second year university students reading physics for a degree; at the same time they would not be within the ordinary vocabulary of non-physics students. Staff members of the Physics Department assisted in the selection of a pool of 70 suitable words. Physics and non-physics students then rated these words according to whether they were familiar with their meaning. Subsequently, on the basis of these ratings, eight critical physics words (CP) were identified which, although highly familiar to physics students, were of minimal familiarity to non-physics students.

Seventy common words (C) were chosen each having a frequency rated greater than 30 in the Thorndike-Lorge Word Count (1944). Each of these was paired, on the basis of word length and initial letter, with one of the 70 physics words. The eight common words paired with CP words are designated CC. The eight critical pair-words were coulomb-column, dipole-dollar, isomer-island, meson-merit, oersted-opinion, rheostat-religion, solenoid-soldier, tetrode-tonight.

The three inspection lists of 60 words contained different proportions of P and C words; the eight critical pairs appearing on each list. Using a table of random numbers the 16 critical words were allocated to the *same* serial positions in all three lists. The remaining 44 positions were filled by words taken at random from the two pools of common and physics words to form three lists as follows:

List X contained 8 CP, 8 CC and 44 P words
List Y contained 8 CP, 8 CC and 44 C words
List Z contained 8 CP, 8 CC, 22 P and 22 C words.

A single test booklet contained the 70 P and their equivalent C words in a randomly determined order. Beside each word was a 5-point scale on which the subject indicated how certain he was that the word was present or absent.

Procedure. Twenty-five physics students and 25 non-physics students were assigned to each of the three inspection list conditions. Each word from the list was projected on to a screen using a Kodak Carousel automatic projector (Model 550R). The exposure time was 3 sec. with a 0·5 sec. interval between words. Subjects were instructed to try to remember the words appearing on the screen. There was a 3 min. interval between the completion of the inspection task and the start of the recognition task. During this time subjects read instructions which were typed on the cover of the test booklet, and which urged them to work steadily without deliberating over any particular word.

Scoring. The 5-point scale in the test booklet enabled subjects to indicate that they were absolutely sure the word was present, almost sure it was present, not sure, almost sure it was absent, and absolutely sure it was absent. Three measures of scoring were used:

(*a*) Number of the 60 List words correctly recognized as having been present (allowing both "absolutely sure was present" and "almost sure was present").

(*b*) Error score on the critical 8 CP and 8 CC words (using the weighting: "not sure" = 1, "almost sure absent" = 2, "absolutely sure absent" = 3).

(*c*) Number of the 140 test words incorrectly recognized (present words rated "almost sure absent" and "absolutely sure absent"; absent words rated "almost sure present" and "abolutely sure present").

RESULTS

Correct recognition of inspection words

The mean numbers of inspection words correctly recognized by the physics and non-physics students are shown in Table I. A 2 × 3 analysis of variance was carried out on these data, and the inspection list was the only significant source of variance ($F(2,144) = 9{\cdot}88$, $p < 0{\cdot}01$).

TABLE I

Mean Number of Inspection List Words Correctly Recognized

Subjects	*Lisi*		
	X	*Y*	*Z*
Physics	47·16	39·12	42·36
Non-physics	46·46	40·12	39·64

Each mean is based upon 25 subjects' responses.

Error scores in recognition of critical words

The means of the weighted scores for errors made in recognizing the 8 CP and the 8 CC words are shown in Table II. An analysis of variance of these data, summarized in Table III, revealed that both the class of critical word (W) and subject group (G) were significant. The effect of inspection list (L) was indicated in the W × L and L × S interactions. To examine significant interactions, t values were calculated. When Lists X and Y were employed, both groups made fewer errors on CP words than on CC words ($p < 0·05$ in each case). However, this difference was not significant for List Z. Physics subjects made fewer errors than non-physics subjects in identifying CP words from List Y and in identifying critical words of both classes from List Z.

TABLE II

Mean Weighted Error Scores

Group	*Physics*			*Non-physics*		
List	X	Y	Z	X	Y	Z
8 CP	3·24	1·80	2·72	4·44	4·76	6·88
8 CC	7·20	8·76	4·20	7·44	7·64	7·16

TABLE III

Analysis of Mean Weighted Error Scores

Source	*d.f.*	*MS*	*F*
Word class (W)	1	718·00	20·9**
List (L)	2	5·52	0·28
Subject group (G)	1	223·33	11·14**
W × L	2	104·70	10·47**
W × G	1	81·01	4·05*
L × G	2	63·44	3·17*
W × L × G..	2	19·37	0·97
Within cell	288	20·03	

* $p < 0·05$.
** $p < 0·01$.

Number of test words incorrectly recognized

This analysis separated two types of errors. A type 1 error occurred when a word present on an inspection list was reported absent; a type 2 error was made when a word not appearing on an inspection list was reported present. The mean number of type 1 and type 2 errors per subject for the 70 physics and 70 common words is shown in Table IV.

TABLE IV

MEAN NUMBER OF ERRORS PER SUBJECT FOR 70 PHYSICS AND 70 COMMON WORDS

		Physics subjects			*Non-physics subjects*		
List	*Error type*	*P Words*	*C Words*	*Total*	*P Words*	*C Words*	*Total*
X	1	7·28	2·01	9·29	8·77	2·01	10·78
	2	0·75	0·89	1·64	2·66	5·83	8·49
Y	1	0·51	14·75	15·26	1·31	13·16	14·47
	2	2·85	3·03	5·88	5·51	4·11	9·62
Z	1	4·06	10·69	14·75	5·69	9·61	15·30
	2	4·43	6·58	11·01	7·33	7·42	14·75

The total of type 1 errors showed little variation between physics and non-physics subjects (9·29 vs. 10·78 for List X; 15·26 vs. 14·47 for List Y; 14·75 vs. 15·30 for List Z). The total of type 2 errors showed greater differences between physics and non-physics subjects (for example, 1·64 vs. 8·49 for List X), and between inspection lists (for example, 1·64 vs. 11·01 for physics subjects). Such comparisons suggest that physics subjects made relatively fewer type 2 errors for each list, and that both groups made least errors for List X, more errors for Y, and most errors for Z.

Comparisons of type 1 errors made to P and C words indicated no large differences between physics and non-physics subjects (for example, for C words: 2·01 vs. 2·01 for List X; 14·75 vs. 13·16 for List Y; 10·69 vs. 9·61 for List Z). Both subject groups made fewer type 1 errors to P words than to C words for List Z and to words in minority class for Lists X and Y. Physics and non-physics subjects were found to differ in the number of type 2 errors made to P and C words. Physics subjects made fewer type 2 errors than non-physics subjects for each of the P and C word comparisons (for example, 0·75 vs. 2·66 and 0·89 vs. 5·83 for List X) and the groups differed also in the proportion of P to C word errors for each inspection list (for example, 0·75 vs. 0·89 and 2·66 vs. 5·83 for List X).

DISCUSSION

There are two findings which are clearly opposed to the expected effects of familiarity and salience. The first, the absence of differences between physics and non-physics students in the number of inspection list words correctly identified (Table I), is contrary to the expectation that familiarity facilitates recognition. The other, that a few common words among many physics words are not recognized better than their paired physics words (Table II, List X), is not in accordance with a von Restorff effect.

Both findings seem to provide supporting evidence for Dale's (1967) conclusion that prior familiarity is not a critical variable when *all* inspection items are presented in a recognition test. It may be argued, for example, that physics students recognize physics words by utilizing "schemata" or organization imposed upon such items during their initial presentation; but that non-physics students rely more heavily upon the array of alternatives in the recognition booklet to supplement imperfectly organized material. Thus physics and non-physics students may employ different strategies during the presentation or retrieval of inspection list material to obtain apparently similar recognition success.

The analysis of types of error (Table IV) provides a means of examining whether different recognition strategies were employed by the two groups. The measure of errors made in reporting "absent" a word appearing on an inspection list (type 1) is a converse index of the amount of inspection materials correctly identified (Table I), and similarly reveals no significant differences between the groups for each list. It is not until errors made to words not on the inspection list (type 2) are considered that important differences between physics and non-physics students become apparent. Physics students were found to make noticeably fewer errors of this type not only to physics words but to common words as well. As the superiority of physics students cannot be attributed solely to their making fewer errors to physics words, it is clear that they benefit from being familiar with the meaning of *all* the inspection list words.

An explanation of the role of familiarity in this experiment is suggested by comparing the different problems encountered in the processing of positive and negative instances. A positive instance occurs when an item in the recognition booklet is identical with a previously presented item from the inspection list; the recognition item is a negative instance if it did not appear in the inspection list. The task of accurately identifying a positive instance only requires that a subject perceives a particular item in the recognition booklet as having appeared as an inspection item. Although physics students were familiar with more of the inspection list items, they were not superior to non-physics students in the identification of positive instances. This is consistent with Dale's (1967) finding that unfamiliar items within each of his lists were just as likely to be correctly recognized as familiar items. The task of accurately identifying a negative instance requires a subject to compare an item in the recognition booklet with his "internal list" of previously experienced inspection items in order to eliminate it correctly. Wason (1959) has stated (in a different context) that "the utilization of negative information would seem to be a relatively mature and sophisticated activity because it implies a store of positive information which can be used as a standard" (p. 103). Hence, the superiority of physics students in identifying negative instances suggests that the scanning of stored information is more efficient when the meaning of all the items in the store is known.

The authors acknowledge the help of Mr. K. F. Taylor for a critical reading of this paper.

REFERENCES

BARTLETT, F. C. (1932). *Remembering*. Cambridge: University Press.

BRUNER, J. S. (1957). On perceptual readiness. *Psychol. Rev.*, **64**, 123–52.

DALE, H. C. A. (1967). Familiarity and free recall. *Quart. J. exp. Psychol.*, **19**, 103–14.

GREEN, D. M. and SWETS, J. A. (1966). *Signal Detection Theory and Psychophysics*. New York: Wiley.

LUCE, R. D. (1959). *Individual Choice Behaviour*. New York: Wiley.

SWETS, J. A. (1964). *Signal Detection and Recognition in Human Observers: Contemporary Readings*. New York: Wiley.

THORNDIKE, E. L., and LORGE, I. (1944). *The Teacher's Word Book of 30,000 Words*. New York: Teachers College, Columbia University, Bureau of Publications.

VON RESTORFF, H. (1933). Uber die Wirkung von Bereichsbildungen im Spurenfeld. *Psychol. Forsch.*, **18**, 299–342.

WASON, P. C. (1959). The processing of positive and negative information. *Quart. J. exp. Psychol.*, **11**, 92–107.

Manuscript received 22nd February, 1968.

RESPONSE-TIMES AT DIFFERENT POSITIONS WITHIN A SENTENCE

BY

R. A. KENNEDY and A. L. WILKES

From the Psychology Department, The University, Dundee

Response latency has often been used to draw inferences about scanning operations in serial lists. Different models have been employed and this experiment tests their predictions when a sentence replaces a serial list. Constituent analysis is used to predict an increase in latency when responding involves crossing a constituent boundary. This prediction is confirmed, and it is found necessary to revise those explanations derived from studies using serial lists to include reference to the syntactic structure of the linguistic material used.

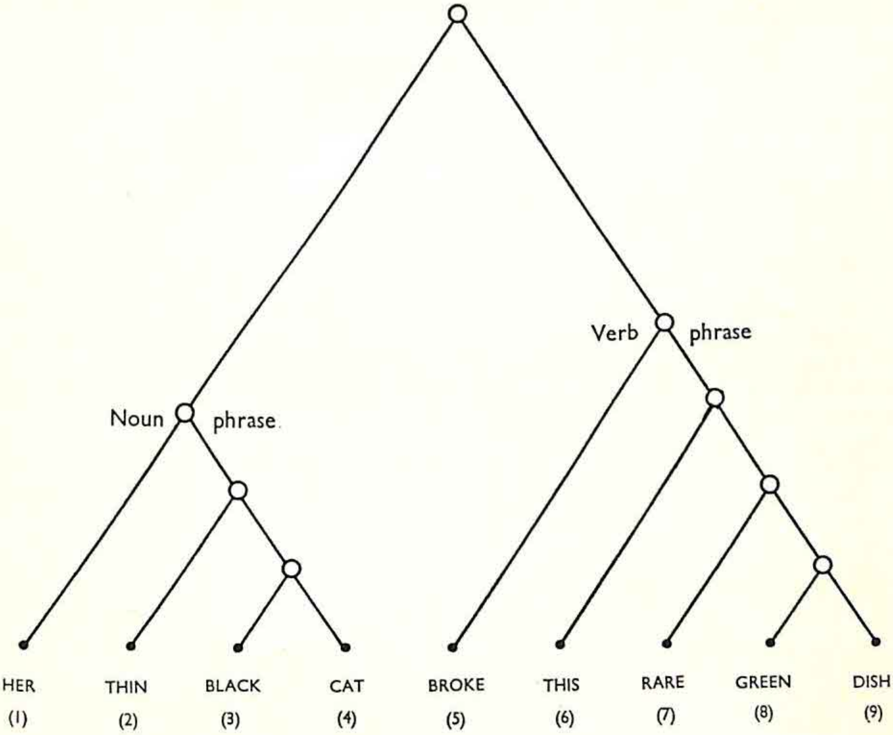
INTRODUCTION

The time taken for a subject to respond to an item in a previously memorized serial list is a useful index of the strategies of storage and retrieval which he employed. Sternberg (1966) presented subjects with a series of short lists, each being followed by a test digit which might or might not have appeared in the list. Subjects were timed while they indicated positive and negative instances by pulling levers. Times to indicate presence or absence of a test digit did not differ significantly, and Sternberg concluded that subjects must engage in an exhaustive serial scan of a list even when a test item was present. As a direct test of this, in a later paper Sternberg (1967) measured the time to respond to an item as a function of its serial position in the list. In a seven-item list response latency varied from about 900 msec. for position 1 to over 1,300 msec. for position 6.

A neglected aspect of information processing theories of verbal learning (e.g. Feigenbaum, 1959; Yntema and Trask, 1963; Hunt, 1963) is that they may be used to generate predictions regarding retrieval time and its relation to serial position in learning and recall. Kennedy and Galloway (1967) using short, easy lists of words confirmed the prediction from Feigenbaum's theory that the time to retrieve items from the centre of a list should be longer than times to retrieve from either the beginning or end. This result, which is in opposition to the findings of Sternberg, was confirmed in a further study (Kennedy, 1968) which tested retrieval time in short serial lists of letters. Items which Feigenbaum termed "perceptually unique anchor points" (i.e. in serial lists of the usual form, the beginning and end), produced relatively shorter response latencies than other positions. The discrepancy between these results and Sternberg's is in all probability related to the fact that they were derived from material which was either very simple or greatly over-learned. It is quite plausible that the scanning of material which has been well learned involves operations which do not reveal themselves when the material is still undergoing a process of consolidation during learning.

In looking beyond serial lists to other forms of ordered material the sentence occupies a position of special interest. A grammatical sentence must meet certain syntactic rules which will determine how particular words group together to form constituents. In scanning and retrieving from within a sentence it follows that considerations are involved, in addition to those applying to simple serial lists. Any analysis must take into account the linguistic structure of the sentence, and in predicting the form of the relationship between serial position and latency a third theoretical possibility must be assessed. This predicts latency differences from the syntactic rules adult subjects follow when perceiving, storing and generating sentences. The most explicit statements of such rules have come from linguists (e.g. Chomsky, 1957), and were originally intended to explain how novel but grammatical sentences could be produced. Thus the original theoretical concern was with linguistic competence; but recently the rule system has come to be used more and more as a guide to the psychological operations involved in verbal performance. An adult subject asked to memorize a sentence and then prompted with single words from it will, if instructed to respond with the following word, probably be influenced by the structural relations within the sentence. Given, for example, the sentence, "Her thin black cat broke this rare green dish," a hierarchical set of constituents may be identified (see Fig. 1). The tree

FIGURE 1



Tree diagram illustrating constituent structure of sentences used. (Example sentence.)

diagram locates a major structural break in the sentence between "cat" and "broke" which suggests that there will be longer latencies when "cat" serves as a prompt word than when, for example, "black" or "broke" do so. Although the constituent analysis illustrated in Figure 1 is based on linguistic rather than psychological principles, a number of recent studies indicate that some borrowing may be justified. In a study of speech perception, Fodor and Bever (1965) used recorded sentences on which clicks had been superimposed and which their subjects had to write down, marking where a click occurred. It was found that clicks objectively preceeding or following major constituent boundaries were likely to be "displaced" resulting in a new location nearer the boundary. From their results the authors argue that the test sentence is divided by the hearer into perceptual units, "on the basis of his knowledge of the constituent structure rules of his language" (p. 418). Similarly, Johnson (1965, 1966), has used immediate constituent analysis to predict the distribution of errors when sentences feature as the responses in paired-associate learning.

Suci, Ammon and Gamlin, (1967) in a similar experiment to that reported here also predicted peak latencies at the noun-verb break. Using children and adults they found a significant F for test-position. A detailed analysis of their results was not provided but latencies appeared to correspond with the constituent structure of the sentences used. Thus from structural criteria alone it can be predicted that in sentences following the syntactic pattern of Figure 1 cue words eliciting responses from position 5 (i.e. across a constituent boundary) will be associated with a longer response latency than words cueing a response from within a constituent. Neither of the scanning models tailored to fit serial lists would place a position of maximum latency quite so early in the sequence and the present experiment has been designed to throw further light on this.

METHOD

Materials. Six sentences consisting of monosyllables and following the syntactic structure illustrated in Figure 1 were used. These were: (*a*) My fat brown dog bit that lame old man. (*b*) Those four good friends made their large blue car. (*c*) That wild young boy rode his big white horse. (*d*) These tall fit men play our long hard game. (*e*) This poor cold girl stole your warm red coat. (*f*) Her thin black cat broke this rare green dish.

Subjects and apparatus. Subjects were three men and five woman: technicians in the Department of Psychology, and members of the secretarial and teaching staff of the University of Dundee. Two tape recorders were used. The first presented the stimulus material, and the second recorded the whole of the experimental session. The sentences for presentation were recorded with the minimum of accentuation by a male voice. The tape from the second recorder was later processed by playing it into a voice key, consisting of an amplifier and trigger with suitable circuitry so that of two successive stimulus inputs, the first would start a timer and the second stop it. Latency between the onset of the stimulus word and the onset of the subject's response was measured: informal reliability checks of the procedure showed an error of measurement not greater than 30 msec.

Procedure. A sentence was presented twice and after each presentation an 8 sec. pause occurred during which the subject repeated it. Each word from the sentence was then presented at 6 sec. intervals in a random sequence. Subjects responded to each, giving as quickly as possible the word immediately following the signal item in the sentence. When the final word of a sentence occurred as a signal, subjects were instructed to say, as quickly as possible, "nothing."

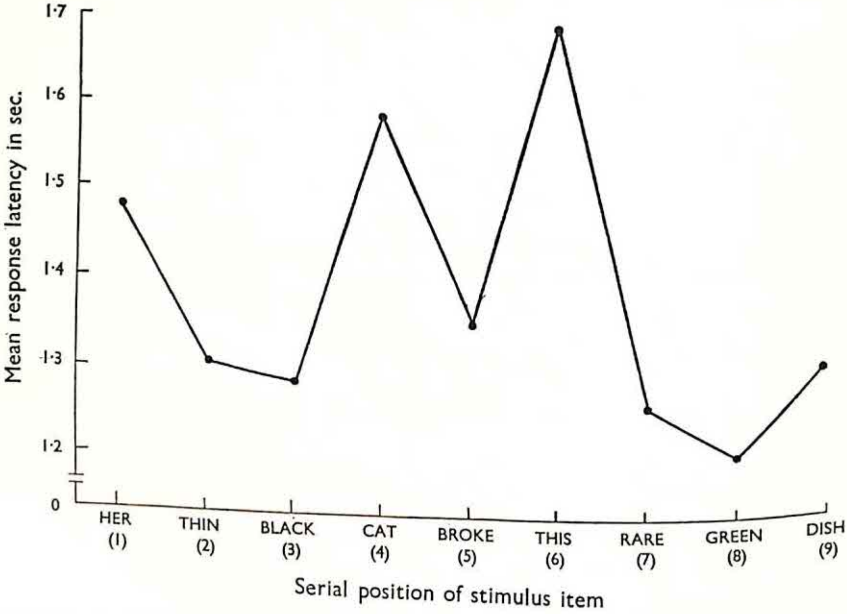
This sequence of two presentations followed by a random test series occurred three times for each of the six sentences. Latencies for all tested positions, including the last, were analysed.

RESULTS

Errors and omissions during recall constituted 6·2 per cent. of the total data for analysis. Following Lindquist (1953, p. 148), these cases were replaced by the subject's mean latency at the relevant test position. A compensating reduction was made in the degrees of freedom for the error term in the analysis of variance.

The mean latency of response to each serial position was averaged for all sentences and is shown in Figure 2.

FIGURE 2



Mean response latency as a function of serial position of stimulus items. (Example sentence.)

The total time to respond to each serial position across all sentences was subjected to an analysis of variance. The obtained distribution of latency scores was tested for skew. The difference between the mean and the median amounted to one-fifth of the standard deviation. The F test is insensitive to skew of this order (Norton, 1952) and transformation of the data seemed unwarranted. Factors in the analysis of variance were the serial position of prompt words and "trials," that is, the three test series for each sentence. Significant effects were position ($F = 12{\cdot}99$, $d.f. = 8, 60$, $p < 0{\cdot}001$), and the interaction of position with "trials" ($F = 4{\cdot}98$, $d.f. = 16, 60$, $p < 0{\cdot}001$). Further interpretation of this interaction was not possible since the effect of position at a given "trial" could well be dependent on the random sequence chosen for the test series. There was no tendency for overall latency to change over the three "trials" (F for "Trials" $= 1{\cdot}31$, $d.f. = 2, 14$).

The Newman–Kuels procedure (Winer, 1962, p. 80) was used to assess the significant differences between particular positions shown in Figure 2. Using a minimal value of significance of $p < 0{\cdot}05$ and adjusting the degrees of freedom to allow for missing cases, significant differences of relevance to the experimental hypotheses were found between positions 3 vs. 4; 4 vs. 5; 5 vs. 6; and 6 vs. 7. More formally stated, it was found position 6 was greater than all other positions but 4; position 4 and 6 greater than all but 1; and 4, 6 and 1 greater than all but 5.

Discussion

The finding of a main effect for serial position makes it appropriate to discuss this in the light of the three theoretical viewpoints examined in the introduction.

The results are clearly incompatible with Sternberg's notion of a serial scan involving an increasing linear relationship of search time with position. If the sentence could be looked on as simply a nine-item array, at the very least it would be expected that response latencies to tests at positions 8 and 9 would be greater than to any other. Sternberg has not extended his argument to deal with anything other than serial lists, but the present findings set a severe limitation on the generality of the scanning model he employs.

The view that the relationship between serial position and latency is curvilinear, that is, that retrieval time for the beginning and end of an array should be shorter than for the centre, gains some limited support. Clearly, anomalous findings on this viewpoint are the latencies found at positions 1, 5 and 9. An increase in latency at position 9 (albeit not significant), may perhaps be accounted for in terms of the nature of the response made, which was a word ("nothing") not in the original display. If a short latency reflects a point of easy access to the array as stored, then position 5 will rank as one of these. This is clearly of relevance in considering the syntactic structure of the sentences used, but poses considerable difficulties for a scanning hypothesis of the kind under consideration.

The comparatively long latency for position 1 is an embarrassment to almost any theoretical account, but it may perhaps be partly explained by reference to sources of interference within the sentence. The importance of controlling for proactive inhibition across sentences in experiments of this sort need hardly be stated. However, control for interference within sentences is particularly difficult given the need for sentences which are both grammatical and consist simply of monosyllables. An enforced duplication of words in certain sentences is found in positions 1 and 6 where there is a need for monosyllabic declarative article or possessive pronoun. In fact only nine words of these grammatical classes are available to fill twelve positions. It should be noted that any compensation for this interference effect that might be employed to bring position 1 to some basal level (i.e. equal to positions 2 and 3), would nevertheless leave position 6 (proportionately reduced) at a significantly higher point than position 1 but below position 4.

The hypothesis that sentence structure is a factor determining ease of access and leading to the prediction that crossing a consituent boundary would be associated with a relatively long latency is obviously supported. Long latency at position 4 contrasts with the short latency found at position 5 and it could be that ready access to the sentence may be made at position 5 in the scanning process. This could imply separate access to the two major constituents of this sentence type but there are reasons for not developing this argument further. The prime difficulty is that of using a predictive model for the generation of sentences when dealing with the processes of storage and retrieval. It seems likely that these operations will share similar principles (see also Mehler and Miller, 1964); the present findings also suggest this may be the case, nonetheless the degree of overlap is still ambiguous.

This research is supported by a grant from the Social Science Research Council.

REFERENCES

CHOMSKY, N. (1957). *Syntactic Structures.* The Hague: Mouton.

FEIGENBAUM, E. A. (1959). *An Information Processing Theory of Verbal Learning.* P–1817, October, Santa Monica, Calif. RAND Corporation.

FODOR, J. A., and BEVER, T. (1965). The psychological reality of linguistic segments. *J. verb. Learning verb. Behav.*, **4**, 414–20.

HUNT, E. B. (1963). Simulation and analytic models of memory. *J. verb. Learn. verb. Behav.*, **2**, 49–59.

JOHNSON, N. F. (1965). The psychological reality of phrase-structure rules. *J. verb. Learn. verb. Behav.*, **4**, 469–75.

JOHNSON, N. F. (1966). On the relationship between sentence structure and the latency in generating the sentence. *J. verb. Learn. verb. Behav.*, **5**, 375–80.

KENNEDY, R. A. (1968). Response latency in the serial learning of short lists. *Brit. J. Psychol.*, **59**, 1–5.

KENNEDY, R. A., and GALLOWAY, A. (1967). Retrieval time and the serial position effect. *Amer. J. Psychol.*, **80**, 430–3.

LINDQUIST, E. F. (1953). *Design and Analysis of Experiments in Psychology and Education.* Boston: Houghton Mifflin.

MEHLER, J., and MILLER, G. (1964). Retroactive interference in the recall of simple sentences. *Brit. J. Psychol.*, **55**, 295–301.

NORTON, D. W. (1952). *An Empirical Investigation of Some Effects of Non-Normality and Heterogeneity on the* F-*Distribution.* Unpublished Ph.D., University of Iowa.

STERNBERG, S. (1966). High speed scanning in human memory. *Science*, **153**, 652–4.

STERNBERG, S. (1967). Retrieval of contextual information from memory. *Psychon. Sci.*, **8**, 55–6.

SUCI, J. G., AMMON, P., and GAMLIN, P. (1967). The validity of the probe-latency technique for assessing structure in language. *Language and Speech*, **10**, 69–80.

WINER, B. J. (1962). *Statistical Procedures in Experimental Design.* New York: McGraw-Hill.

YNTEMA, D. B., and TRASK, F. P. (1963). Recall as a search process. *J. verb. Learn. verb. Behav.*, **2**, 65–74.

Manuscript received 30th November, 1967.

LETTER-NAMING TIME AS A FUNCTION OF SET FAMILIARITY AND SYMBOL DISTINCTIVENESS

BY

W. A. HERSHBERGER, P. R. TRANTINA and KATHY COSGROVE

From Northern Illinois University

Two symbol-naming experiments were conducted assessing the dependence of Fitts and Switzer's set-familiarity effect upon symbol distinctiveness. Sixty college males named printed letters presented in a strobotron tachistoscope, the letter always being selected from a preannounced set of three. A voice key detected the response. Experiment I found the Fitts and Switzer finding to be a joint effect: response latency for naming the symbol B in the unfamiliar but distinctive set, VBO, was intermediate to that for the familiar distinctive set, ABC, and the unfamiliar, homolographic set, PBE, the two sets used by Fitts and Switzer. Experiment II, a factorial combination of set familiarity and symbol distinctiveness yielded a significant interaction such that with homolographic symbols, set familiarity was associated with an increase in reaction time. The results were interpreted as consistent with an hypothesis that the set-familiarity effect relates to symbol-identification time as opposed to response-identification time.

INTRODUCTION

Fitts and Switzer (1962) have found that the response latency for naming the visual symbol B is less when it appears alternately with the symbol A and C than when it appears alternately with the symbols E and P. They attributed this effect to the difference in familiarity of the two sets of letters: *A*, *B*, and *C* vs. *P*, *B*, and *E*. However, their finding might just as readily be interpreted as revealing the effect of stimulus similarity: the symbol B is graphically more distinctive in the set ABC than in the set BPE where the symbols P and E comprise mere subsets of the symbol B. Or, since the two interpretations do not appear mutually exclusive, the Fitts and Switzer finding may reflect a joint effect. It was the purpose of the present research to assess the relative influence of the two factors.

Naming alpha-numeric characters is a perfunctory task well routinized by most literate individuals. The verbal symbol or name of each letter category is so firmly associated with its corresponding visual symbol and so fully dissociated from all others that the task involves no apparent deliberation or choice. Accordingly, it is sometimes found that the response latency for naming individual alpha-numeric symbols is independent of the number of equiprobable characters from which the symbol is selected (Brainard, Irby, Fitts, and Alluisi, 1962; Morin and Forrin, 1962; Mowbray, 1960). However, where speed is of the essence, an experimental subject may be pressed into naming an alpha-numeric symbol before he has fully inspected it, that is, before the symbol, let alone the letter category it represents, has been thoroughly identified. In this case, the naming of the symbol becomes, clearly, a disjunctive reaction with the subject deciding with limited stimulus information which of several symbols he is viewing.

If he is to reduce his response latency in this way he must, however, possess certain prior information which allows him to "deduce" the identity of the symbol from incomplete—and otherwise indeterminate—stimulus information. Such *a priori* information may be provided him by an announced reduction in the number of symbols he is called to identify. For example, if he is advised that only three symbols from the 26-character alphabet are to be presented, say, A, B and C, he need only determine for each stimulus presentation whether the symbol incorporates curved or straight lines: if there are no curved lines the symbol is an "A"; if there are no straight lines it is a "C"; if there are both straight and curved lines it is a "B". The more distinctive the symbols the less stimulus information required to distinguish between them.

Providing it is sufficiently simple to employ, such a deductive technique may allow the subject to reduce his response latencies by some small but measurable amount (cf. Rapoport, 1959). Not surprisingly, several investigators (Forrin and Morin, 1966; Morin, Konick, Troxell and McPherson, 1965) have recently reported that the response latency for naming visual, alpha-numeric symbols presented individually may be reduced by limiting the number of alternative symbols presented to a total of approximately four or less. It is doubtful if a limited set-size is the only prerequisite for this effect. The symbols themselves must, presumably, differ from one another sufficiently to allow the subject to choose between them with very little stimulus information. Also, since the subject must remember the particular alpha-numeric characters which comprise the set of symbols to be named, familiarity with those symbols as a set should facilitate his use of the deductive, symbol-identification process. It is this last factor, familiarity of symbol set, to which Fitts and Switzer attribute their findings that the symbol B is named more rapidly in the set ABC than in the set BPE. However, they did not control for the distinctiveness of the B in their two experimental sets. The experiments reported here were designed to do this.

In the first experiment, three sets of symbols were used, ABC, VBO, and PBE, the set VBO being as unfamiliar as BPE but including symbols roughly as distinctive from one another as those in the familiar set ABC. It was found that the response latency for naming the symbol B was shortest for the set ABC, next shortest for the set VBO, and longest for the set BPE. In other words, both stimulus distinctiveness and set familiarity contribute to the specific finding reported by Fitts and Switzer. However, it was impossible to determine to what extent the two factors operate independently of each other since the familiar-set comparison involved only distinctive sets whereas the stimulus-similarity comparison involved only unfamiliar sets. To explore this issue, a second experiment was conducted: a multidimensional factorial design combined each of two levels of stimulus similarity with each of two levels of set familiarity.

METHOD

Subjects. Sixty male students attending Northern Illinois University served as subjects. Twelve served in Experiment I. Forty-eight served in Experiment II.

Apparatus. The symbols to be named were black capital letters printed in the centre of white paper cards, four inches square. The cards were presented in a Lafayette, strobotron tachistoscope and were viewed from a distance of approximately 20 in. A microphone mounted on a small boom was placed near the subject's mouth and operated an electronic voice key. Response latencies were measured with a Hunter Klockounter, accurate to 0·001 sec. The experimenter activated the tachistoscope and the Klockounter simultaneously by throwing a single DPDT toggle switch; the subject's voice stopped both instruments simultaneously by activating the voice key.

Procedure. The subjects were tested individually. Each student was instructed that on each trial he was to name as rapidly as possible the letter which would become illuminated in the viewer (tachistoscope) before him. Speed rather than accuracy was stressed. If he misidentified a symbol, that symbol was presented again at the end of that particular block of trials. Nevertheless, errors were infrequent.

An experimental session comprised several blocks of 75 trials each. During any one block of trials only three different letters were presented. They were presented equally often and in a random order. At the beginning of each block of 75 trials, the subject was always shown the set of three letters he would be required to name.

Each trial was prefaced with a signal of "ready" approximately 1 sec. before the symbol was illuminated. The intertrial interval was approximately 8 sec. and the inter-block interval was of the order of several minutes.

To familiarize him with the procedure, each subject was given nine practice trials naming the colour of a small square centrally located on each of three practice cards (red, yellow, green) each presented three times in a random order.

Experiment I. Capital letters printed in Berling Italics, 48 pt., were used (Letraset No. 317). There were three blocks of trials, one block with the distinctive symbols comprising the familiar set ABC, one block with the distinctive symbols comprising the unfamiliar set VBO, and one block with the homolographic symbols comprising the unfamiliar set PBE. Each of the six different permutations of these three blocks was randomly assigned to a different set of two subjects.

Experiment II. A multidimensional design was employed. Two blocks of trials were used, one block with a familiar set of symbols DOG (or GOD), and the other block with an unfamiliar set of symbols GCD. Half the subjects were given the familiar set first, and half the unfamiliar. Half of each of these groups were told that the letters of the familiar set comprised the word "GOD"; the other half were told that the set comprised the word "DOG." Half of each of these subgroups were presented relatively distinctive symbols in the form of capital letters printed in Berling Italics, 48 pt. Letraset No. 317). The other half was presented homolographic symbols comprised basically of a circle $\frac{13}{16}$ in. in diameter. The O was a full circle. The vertical upright of the symbol D was $\frac{1}{4}$ in. in length. The gap in the symbols C and G was, in each case, $\frac{1}{8}$ in. The G included a $\frac{1}{8}$ in. vertical line beginning at the lower edge of the gap and extending downward. The symbols were drawn in black ink with a ball-point pen. Although they could be distinguished, the symbols were not readily discriminable from one another.

Including the differences in the letters themselves, Experiment II comprised a four-factor design: Letters, Set-Familiarity, Symbol Distinctiveness, and Letter Order (GOD vs. DOG in the familiar set). The design was mixed, with all values of the first two factors administered to each subject, and each combination of values of the last two factors administered to a different, randomly-selected quarter of the subject pool ($N = 48$).

RESULTS

Experiment I. The overall mean response latency for naming the symbol B, the one letter which appeared in all three trial blocks, was 0·431 sec. in set ABC, 0·456 sec. in set VBO and 0·472 sec. in set PBE. All differences among these means are significant (Sign Test: $p < 0{\cdot}05$).

Experiment II. Mean response latencies for naming the symbols D and G (the two letters which appeared in both trial blocks) are shown in Table I as a function of Symbol Distinctiveness, Set Familiarity and Letter Order (GOD vs. DOG).

TABLE I

MEAN RESPONSE LATENCIES TO THE SYMBOLS D AND G AS A FUNCTION OF STIMULUS DISTINCTIVENESS, SET FAMILIARITY AND LETTER ORDER

(Data in sec.)

	Familiar set		*Unfamiliar set*	
Stimulus distinctiveness	*Letter G*	*Letter D*	*Letter G*	*Letter D*
Distinctive:				
Letter order:				
G–D	0·492	0·484	0·518	0·503
D–G	0·482	0·487	0·503	0·501
Homolographic:				
Letter order:				
G–D	0·619	0·623	0·618	0·581
D–G	0·613	0·601	0·607	0·594

A four-way analysis of variance of the subjects' mean response latencies yielded four significant effects: Symbol Distinctiveness, Letter (G vs. D), a first-order interaction between Set Familiarity (GOD vs. GCD) and Letter (G vs. D), and finally an interaction between Set Familiarity and Symbol Distinctiveness. The mean response latency was shorter for the distinctive symbols, 0·496 sec., than for the homolographic symbols, 0·607 sec. ($F = 53{\cdot}93$, $d.f. = 1/44$, $p < 0{\cdot}001$). The mean response latency to the symbol D, 0·547 sec., was shorter than the mean response latency to the symbol G, 0·557 sec. ($F = 5{\cdot}17$, $d.f. = 1/44$, $p < 0{\cdot}05$). The interaction of Set Familiarity and Letter ($F = 7{\cdot}67$, $d.f. = 1/44$, $p < 0{\cdot}01$) was such that the letter G was more readily identified than the letter D in the familiar set, 0·549 sec. vs. 0·552 sec., whereas G was less readily identified than the letter D in the unfamiliar set, 0·562 sec. vs. 0·545 sec. The interaction of Set Familiarity with Symbol Distinctiveness ($F = 7{\cdot}39$, $d.f. = 1/44$, $p < 0{\cdot}01$) was such that for distinctive symbols, mean response latency was shorter for the familiar set, 0·486 sec., than for the unfamiliar set, 0·507 sec., whereas for homolographic symbols the opposite was true: the mean response latency was shorter for the unfamiliar set, 0·599 sec., than for the familiar set, 0·614 sec.

Discussion

The present results show that although Fitts and Switzer's set-familiarity effect is genuine, the nature of the effect somehow depends upon the distinctiveness of the symbols being named: when the symbols are distinctive, familiarity with the symbol set reduces reaction time, as reported by Fitts and Switzer; however, when the symbols are homolographic and difficult to distinguish, set-familiarity may be associated with an increase in reaction time, either as a causative factor or as the accidental result of being confounded with a more significant variable which does vary as a function of symbol distinctiveness. The relative merits of these two interpretations are discussed a little later.

Although the familiar-set effect appears to depend upon symbol distinctiveness, the converse is not true. Distinctive symbols are named more rapidly than homolographic ones, irrespective of the familiarity of the set of symbols named (cf. Crossman, 1955).

The present finding of a shorter response latency for the letter D than for the letter G may be attributed to a difference in their phonetic labels. "D" is a harder phoneme better suited for detection by the voice key. Hence, this difference, also noted by Fitts and Switzer, appears trivial.

What does not appear trivial, however, is the interaction of the variables Letter and Set Familiarity, for it illustrates further the significant effect of symbol distinctiveness upon symbol identification. The letter G was more readily identified in the familiar set, GOD (or DOG), where it was the only open figure than in the unfamiliar set, GCD, where it was one of two open figures. Conversely, the letter D was more readily identified in the unfamiliar set, GCD, where it was the only closed figure than in the familiar set, GOD (or DOG), where it was one of two closed figures. Not only are the symbols of a distinctive set more rapidly identified than those of a homolographic set, but the more distinctive or unique a symbol is within its own set, the more rapidly is it identified, relative to the other members of its set.

Apparently the information prerequisite to accurate symbol naming of the type investigated here depends not so much upon the number of verbal responses (letter categories) involved, but upon the number of alternate groups of homolographic visual symbols comprising the stimuli being identified. The subject appears to scan mentally the letters from which he is to select his response not as individually printed symbols but as members of a hierarchy of generic sets and subsets of symbols, the subsets at each successive level of the hierarchy being composed of symbols which are relatively more homolographic than those of previous levels. It is as if he processes his stimulus information through a sequence of mental sorters with each sorter assigning the input to a progressively finer homolographic category until only one letter remains. Symbol identification time depends evidently upon the number of sorts required. Reducing the number of symbols to be identified reduces the number of sorts required, particularly when (*a*) the symbols are distinctive from each other, and (*b*) the symbols comprise a set easily remembered, thereby preventing irrelevant sorting.

However, why then should set familiarity ever increase naming times as it appears to have done here in the case of the homolographic symbols? Presumably it should not. Plausibly it did not. The significant interaction appears to be the effect of yet another instance of differential symbol distinctiveness: as symbols to be discriminated become increasingly homolographic, the distinctive characteristic of each becomes specific to a smaller and smaller portion of the symbol itself. In discriminating O and Q, for instance, only the lower, right-hand portion of the symbols provide any discriminative information. In general, the more homolographic a set of symbols become, the greater the need to scan each symbol for discriminative stimuli. Further, as scanning time becomes an increasingly significant factor, so also do the exact configurations of the letters within a set become increasingly critical. Consider the homolographic symbols used in the present study which were comprised basically of a circle. The symbols G, C, and D could be discriminated solely in terms of their right-hand contours, whereas this was not possible with the symbols G, O, and D. In the latter case, O differed from D only in terms of its left-hand contour while differing from G only in terms of its right-hand contour. Presumably in naming letters from the familiar set, GOD, the subjects frequently scanned the symbol twice, once for discriminative stimuli distinguishing the symbol categories "O, D" vs. "G" and again for stimuli distinguishing the categories "O" vs. "D" (or, "G, O" vs. "D", and then "G" vs. "O"). In naming the letters of the unfamiliar set, GCD, the subjects needed only to scan the symbol for its right-hand contour which provided the discriminative stimulus in every case. Apparently, the increased scanning required by the set GOD more than compensated for any advantage that accrued to it by virtue of its familiarity. Unfortunately, the absolute magnitude of the two effects cannot be ascertained since they are perfectly confounded. And as a consequence, we cannot actually tell whether there is a direct interaction between symbol distinctiveness and set-familiarity per se. We may only note that as a set of symbols to be named becomes increasingly homolographic, the topographical character of the symbols becomes increasingly significant, with the potential of obscuring any effects attributable to set familiarity.

That the process of symbol identification investigated here involves stimulus discrimination as opposed to response differentiation is evidenced further by the absence of an interaction involving the factors Letter (G vs. D) and Letter Order (G–D vs. D–G). If verbal response categories were being scanned, one could expect the response latency to the symbol D to be shorter for the set D–G than for the set G–D, and the latency to the symbol G to be shorter for the set G–D than for the set D–G. However, no interaction involving these two factors proved to be statistically significant ($p = 0{\cdot}05$). Of course if the scanning of response categories were exhaustive as suggested by Rapoport (1959) and Sternberg (1966) then no such interaction would be expected, but neither would one have expected the obtained interaction of Letter and Set Familiarity. Clearly, the present results imply that symbol naming of the type investigated here involves scanning one's repertoire of visual-symbol categories rather than verbal-responses categories.

The work reported herein was supported in part by a grant from the U.S. Office of Education, Department of Health, Education, and Welfare. The authors thank Harold Taylor for preliminary work on the project.

REFERENCES

BRAINARD, R. W., IRBY, T. S., FITTS, P. M., and ALLUISI, E. A. (1962). Some variables influencing the rate of gain of information. *J. exp. Psychol.*, **63**, 105–10.

CROSSMAN, E. R. F. W. (1955). The measurement of discriminability. *Quart. J. exp. Psychol.*, **7**, 176–95.

FITTS, P. M., and SWITZER, G. (1962). Cognitive aspects of information processing: I The familiarity of S–R sets and subsets. *J. exp. Psychol.*, **63**, 321–9.

FORRIN, B., and MORIN, R. E. (1966). Effect of contextual associations upon selective reaction time in a numeral-naming task. *J. exp. Psychol.*, **71**, 40–6.

MORIN, R. E., and FORRIN, B. (1962). Mixing of two types of S–R associations in a choice reaction time task. *J. exp. Psychol.*, **64**, 137–41.

MORIN, R. E., KONICK, A., TROXELL, N., and MCPHERSON, S. (1965). Information and reaction time for "naming" responses. *J. exp. Psychol.*, **70**, 309–14.

MOWBRAY, G. H. (1960). Choice reaction time for skilled responses. *Quart. J. exp. Psychol.*, **12**, 193–202.

RAPOPORT, A. (1959). A study of disjunctive reaction times. *Behav. Sci.*, **4**, 299–315.

STERNBERG, S. (1966). High-speed scanning in human memory. *Science*, **153**, 652–4.

Manuscript received 20th February, 1968.

THE EFFECT OF ELECTRIC SHOCK ON RATS' CHOICE BETWEEN FAMILIAR AND UNFAMILIAR MAZE ARMS: A REPLICATION

BY

M. H. SHELDON*

From the Department of Psychology, University of Hull

It is argued that experiments showing that noise and electric shock produce gross changes in rats' locomotor activity are only weak evidence for a relationship between fear and exploration. Stronger evidence comes from an experiment in which fewer shocked than unshocked rats chose the unfamiliar arm of a maze. Certain details of this experiment are unclear, and an attempt is made here to confirm its findings. In one experiment, shocked rats again chose the unfamiliar arm less often than unshocked rats, and in a second experiment it was shown that this finding cannot be explained by saying that shocked rats choose at random. These results are discussed in relation to recent suggestions about exploration in rats.

INTRODUCTION

It is quite commonly suggested that there is a relationship between exploratory behaviour and fear, and a number of experimenters (Broadhurst (1957) and Montgomery and Monkman (1955) with rats; Baron (1964) with mice) have investigated the effects on locomotor activity of stimulation assumed to produce fear (electric shock, and loud noises). But to show that a treatment affects locomotor activity is not necessarily to show that it affects exploration, if by exploration we mean behaviour under the control of stimuli arising from the environment. The case for a connection between fear and exploration would be stronger if it could be shown that in addition to changes in activity it produced changes in the pattern of an animal's response to the objects in its environment.

In view of this, an experiment by Thompson and Higgins (1958) may be important. They showed that rats who had just been shocked chose to enter the more familiar of two maze arms in a situation where unshocked rats chose the less familiar arm. Some details of the procedure are not clear in the published account. Both the duration ("1–3 sec.") and intensity ("approximately 1·0–1·7 mps.") of shock seem to have been quite variable. The authors say that at the end of the 10 min. exposure period rats "were moved from the habituation compartment and immediately placed in the starting-box with the door closed." But what happened if at the end of 10 min. the rat was already in the starting-box: was it taken out and replaced, or was it left there?

In the two experiments presented here I have tried to confirm Thompson and Higgins' results using a similar piece of apparatus and describing the procedure more fully.

EXPERIMENT I

Subjects

Forty-three hooded rats (24 female and 19 male) were used, bred in the laboratory; they were about 100 days old. All rats had been tested in an open field for 3 min. on each of 4 successive days shortly before taking part in this experiment.

Apparatus

Rats were tested in a Y-maze in which the arms were set at equal angles. One arm (referred to as the start-box) was 5 in. long and had a floor of stainless steel bars running across its width. The other two arms (referred to as the maze arms) were 10 in. long. They and the triangular choice area at the junction of the three arms had a plain wooden floor. All three arms were 6 in. wide and 12 in. high. The maze was made throughout of unpainted hardboard. Entry to each arm was controlled by a guillotine door. The maze was placed approximately in the centre of a large dark cupboard. A 15 watt bulb hung 2 ft. above the centre of the maze, and white noise of low intensity was delivered by a loudspeaker standing behind the start-box. When a morse key was pressed, shock lasting 0·5 sec. was delivered through the grid floor; the current, assuming a resistance of 10,000 ohms, was 0·8 mA.

---

*Now at Department of Psychology, University of Reading.

Procedure

Each rat was carried from its living cage into the cupboard, and put into the maze at the end of one of the two arms of the Y, facing away from the choice point. The guillotine doors at the entrance to that arm, and at the entrance to the start-box, were open, and the rat was free to move between the arm and the start-box for 10 min. The door leading to the second arm was closed throughout. After 10 min. the rat was taken out of the maze (wherever it was) and placed (or replaced) in the start-box with the door closed. The start-box was covered with a roof of transparent "Perspex," and after 5 sec. half of the rats were given a single shock. Five sec. later (or, for unshocked rats, 10 sec. after the roof had been placed on the start-box) all three maze doors were opened, and the rat was allowed to choose between the two arms of the Y. It was regarded as having made its choice as soon as all its feet had crossed the threshold of one of the arms. If a rat failed to choose within 15 min. it was taken out of the maze. As far as possible equal numbers of rats (both males and females) were introduced into each of the two maze arms, and equal numbers of rats introduced into each arm were shocked and not shocked. Trials for shocked and unshocked animals followed one another haphazardly.

Results

Table 1 shows how many shocked and unshocked rats chose the arm to which they had not been exposed previously (the "unfamiliar" arm) and how many chose the arm to which they had been exposed (the "familiar" arm). Results for males and females are shown separately. One animal—a shocked female—failed to choose within 15 min. Since the hypothesis being tested was that shocked rats would choose the familiar arm more often than unshocked rats, this rat has been treated as though it chose the *unfamiliar* arm, and its score appears in Table 1. Combining the results from males and females we find that more shocked than unshocked animals chose the familiar arm ($\chi^2 = 8{\cdot}43$, $d.f. = 1$, $p < 0{\cdot}01$).

TABLE I

		Shocked	*Unshocked*
Number of rats entering familiar arm	Male	6	0
	Female	5	1
	Total	11	1
Number of rats entering unfamiliar arm	Male	4	9
	Female	7	11
	Total	11	20

Discussion

Fewer shocked rats entered the unfamiliar maze arm, and in this the results agree with those of Thompson and Higgins. But Thompson and Higgins showed that shocked rats *preferred* the familiar arm (24 animals out of 30) whereas the present result could be explained by assuming that shock has such a disruptive effect on behaviour that the choices of the shocked animals were random.

If Thompson and Higgins' result is repeatable it would be natural to assume that the probability of a rat's choosing the familiar arm would increase with the severity of the shock. The shock used here seems to have been both shorter and less intense than that used by Thompson and Higgins, and it might, therefore, be expected that if the experiment were repeated, using a more severe shock, shocked rats would show a significant preference for the familiar arm. There is, however, a practical difficulty. Even in the present experiment one rat failed to choose within 15 min., and this, together with common observation of shocked rats, suggests that if the shock were made more severe a substantial number of rats would fail to leave the start-box within a reasonable time. The problem would then arise of how these rats were to be treated. It would clearly not do simply to ignore them: it could always be argued that rats who are very slow to choose are precisely the ones who, when they eventually do choose, enter the unfamiliar arm. Since it is hard to see any legitimate basis for excluding an equal number of rats from the unshocked group, the only satisfactory course would be to assume that these rats would have chosen the unfamiliar arm (contrary to the prediction). But as soon as the number failing to choose becomes substantial, even if those rats who do choose all choose the familiar arm, a rather large number of rats would be needed to show an effect.

EXPERIMENT 2

If we are concerned simply to show that the results of Experiment 1 cannot be explained by saying that choice following shock is random, then matters are less difficult. If we can show, using a more intense shock, that of those rats who do choose within a reasonable time, a significantly large number choose the familiar arm, we can at least claim that shock does not as a general rule disorganize choice behaviour. It would then be more reasonable to interpret Experiment 1 as showing that shocked rats increase their preference for the familiar arm. In an experiment with this limited aim we should be entitled to ignore the question of how any rats that did not choose would eventually have chosen.

Subjects

Forty-five hooded rats (20 male and 25 female) were used. They were of the same stock as those used in Experiment 1, and had had similar experimental experience.

Apparatus

This was the same as for Experiment 1. The intensity of the shock was increased to 1·7 mA. (assuming a resistance of 10,000 ohms).

Procedure

As in Experiment 1, rats were exposed to one maze arm and the start-box for 10 min., and then placed in the start-box with the door closed and a "Perspex" roof in place. Five sec. later all rats were given two 0·5 sec. shocks (1·7 mA.) in rapid succession. Five sec. after the second shock all three doors were opened and the rat was allowed to choose. Any rat who had not made a choice was given a third shock (0·5 sec., 1·7 mA.) precisely 5 min. after the doors had been opened.

Results

Table II gives the choice data. It shows results separately for males and females, and for rats taking more or less than 5 min.

TABLE II

		Rats choosing within 5 min.	*Rats shocked after 5 min.*
Number of rats entering familiar arm	Male	8	5
	Female	14	5
	Total	22	10
Number of rats entering unfamiliar arm	Male	7	0
	Female	5	1
	Total	12	1

If the effect of shock is simply to disrupt choice behaviour so that it is random, equal numbers of rats would be expected to choose the familiar and the unfamiliar arms. Looking first at those rats who choose within 5 min., and combining results from males and females, we find that the number choosing the familiar arm is scarcely significant ($\chi^2 = 2{\cdot}94$, $d.f. = 1$, $p<0{\cdot}10>0{\cdot}05$). Taking females alone we find that the effect is greater ($\chi^2 = 4{\cdot}26$, $d.f. = 1$, $p<0{\cdot}05$). Looking next at rats shocked at the end of 5 min., we find that a significantly large number choose the familiar arm (by the Binominal test, $p = 0{\cdot}006$).

Discussion

These results make it more difficult to argue that the shocked rats in Experiment 1 were choosing unsystematically: it seems that choice shortly after shock need not be random. Of rats choosing within 5 min., the females at least showed a clear preference for the familiar arm. The results from rats given a third shock at the end of 5 min. are particularly striking. At the end of 5 min. all these animals were still in the start-box, looking out through the open door; they entered one of the maze arms as soon as the third shock was given, and although under these circumstances their behaviour might have been expected to be disorganized, 10 out of 11 choose the familiar arm. It cannot, of course be assumed that these rats would eventually have chosen the familiar arm if they had been left unlimited time for choice after the initial pair of shocks, and for this reason we must be cautious about combining their results with those from the other animals. But if we simply ask the general question "Need rats who have just received a shock choose at random?" it seems to be legitimate to combine the data. In this case choice of the familiar arm occurred at a high level: $\chi^2 = 8{\cdot}02$, $d.f. = 1$, $p<0{\cdot}01$.

GENERAL DISCUSSION

Taken together, the results of these experiments confirm the findings of Thompson and Higgins. Shocked rats are less likely than unshocked rats to choose an unfamiliar maze arm, and it seems that this difference has to be interpreted as an increased preference for the familiar arm.

The experiments referred to in the Introduction seem to have been concerned with the idea that an animal's exploratory behaviour will bear some relationship to its general level of fear or emotionality; a treatment like shock (delivered inside or outside the testing situation) may be expected to affect exploration if it increases the animal's fear. This view of a relationship between fear and exploration should perhaps be distinguished from that put forward by Mowrer (1960) and by Halliday (1967). According to this, animals approach features of the environment about which they are moderately fearful, although of course they will avoid those that arouse intense fear. These two notions about the effect of fear are not, of course, contradictory, and it is tempting to see in the present experiments a point at which they might be brought together. If we assume that the amount of fear aroused by a particular feature of the environment, depends partly on the animal's general level of fear, and that this may be raised by shock, then, given the further assumption that an animal's tendency to approach a stimulus first increases and then decreases with the amount of fear it elicits, our results can perhaps be explained. We might argue that unshocked rats enter the unfamiliar arm because it makes them moderately fearful, whereas for the shocked rats, whose general level of fear has been artificially raised, the unfamiliar arm arouses such intense fear that they avoid it.

If this view were correct it might be expected that the tendency to choose the familiar arm after shock would be even more marked in animals whose behaviour in other situations led us to call them relatively fearful or emotional. Eysenck and Broadhurst (1964) summarize a large amount of work on the Maudsley strains of emotionally reactive and nonreactive rats, showing that defecation scores in an open field are correlated with behaviour in a number of situations where we might expect differences in emotional responsiveness to be important. The rats used in both the present experiments had previously been tested for 4 days in an open field, and their defecation scores (total number of boluses) were available. It might be expected that the defecation scores of shocked rats who chose the familiar arm would be higher than those of shocked rats who chose the unfamiliar arm. But this is not so. In Experiment 1 shocked rats choosing the familiar arm had significantly *lower* defecation scores ($U = 19$, $p<0{\cdot}02$, two-tailed). In Experiment 2 (taking only those rats who choose within 5 min.) the difference is smaller but in the same direction (by the U-test, $p<0{\cdot}10$, two-tailed).

How damaging these findings are for the explanation sketched above depends partly, of course, on how generally it is true that defecation scores measure underlying differences in emotional responsiveness. What holds for Maudsley reactive and nonreactive rats may not hold for other strains. Thus, for example, in a number of experiments by Denenberg and his associates (Denenberg and Morton, 1964; Denenberg, Morton, and Haltmeyer, 1964 (Expt. I); Denenberg and Grota, 1964) on the effects of rearing conditions, various behavioural differences that it has seemed reasonable to attribute to differences in emotionality were *unrelated* to defecation scores. The absence of differences in defecation in the present experiments, then, would not be too serious. But it is difficult to explain why there should have been differences opposite to those that might have been expected.

I am most grateful to Dr. D. C. Kendrick for making available to me the open-field records of the rats used in these experiments. I should also like to thank Mr. R. F. Drewett for some most valuable assistance.

REFERENCES

BARON, A. (1964). Suppression of exploratory behavior by aversive stimulation. *J. comp. physiol. Psychol.*, **53**, 299–301.

BROADHURST, P. L. (1957). Determinants of emotionality in the rat I. Situational factors. *Brit. J. Psychol.*, **48**, 1–12.

DENENBERG, V. H., and GROTA, L. J. (1964). Social-seeking and novelty-seeking behaviour as a function of differential rearing histories. *J. abnorm. soc. Psychol.*, **69**, 453–6.

DENENBERG, V. H., and MORTON, J. R. C. (1964). Infantile stimulation, prepubertal sexual-social interaction, and emotionality. *Anim. Behav.*, **12**, 11–3.

DENENBERG, V. H., MORTON, J. R. C., and HALTMEYER, G. C. (1964). Effects of social groupings upon emotional behaviour. *Anim. Behav.*, **12**, 205–8.

EYSENCK, H. J., and BROADHURST, P. L. (1964). Chapter 20 in EYSENCK, H. J. (Ed.), *Experiments in Motivation*, Oxford: Pergamon Press.

HALLIDAY, M. S. (1967). Exploratory behaviour in elevated and enclosed mazes. *Quart. J. exp. Psychol.*, **19**, 254–63.

MONTGOMERY, K. C., and MONKMAN, J. A. (1955). The relation between fear and exploratory behaviour. *J. comp. physiol. Psychol.*, **48**, 132–6.

MOWRER, O. H. (1960). *Learning Theory and Behaviour.* New York: John Wiley.

THOMPSON, W. R., and HIGGINS, W. H. (1958). Emotion and organized behaviour: experimental data bearing on the Leeper-Young controversy. *Canad. J. Psychol.*, **12**, 61–8.

Manuscript received 1st June, 1968.

EFFECT OF SIMILARITY OF INTERPOLATED MATERIALS ON SHORT-TERM RECALL

BY

V. K. KOTHURKAR

From the University of Poona

Two experimental studies bear out the assertion that the probability of short-term recall of verbal units is adversely affected by the similarity of interpolated items. The possibility is conceded that some confounding of trace decay and interference factors may occur whenever an interval between an original presentation and its subsequent recall is filled with some intervening activity; yet it may be concluded that the results broadly corroborate and support the interference hypothesis as the major explanation of short-term forgetting, and of the decisive role of similarity.

EXPERIMENT I

Method

Design of the experiment. The basic paradigm of the Peterson and Peterson (1959) short-term memory experiment was adapted into a 4 × 3 factorial design. The four different kinds of interpolated activities filling the retention intervals were spelling out CVC nonsense syllables, reading three-digit random numbers, perceptual discrimination—a simple perceptual task of judging and reporting which of the two small presented lines was longer—and colour naming. These four kinds of materials were presented, during the retention interval, on a memory drum at a constant rate of 1·5 sec. per frame.

The above-mentioned four different kinds of intervening activities were systematically combined with three retention intervals of 6, 15, and 24 sec. yielding in all 12 variations of experimental treatments.

Materials. Forty-eight three-consonant units with Witmer association value no greater than 25 per cent. constituted the central pool of to-be-remembered materials. Random samples of 24 such consonant trigrams from the common pool constituted the stimulus items for each of the 12 experimental treatments. Other filling materials were CVC nonsense syllables of low association value, three-digit random numbers, lines drawn in black ink and of varying lengths and some coloured papers.

The similarity between the original and interpolated materials, however, was not systematically varied. Although the similarity–dissimilarity gradient was not explicitly worked out, it was presumed that the intervening CVC materials were most similar to the original consonant trigrams, and that the three digit numbers came next as having a certain formal likeness with the three consonant units. The similarity of the two remaining perceptual tasks with the original learning task was also not actually ascertained but was assumed—not unreasonably, it is believed—to be rather low on the resulting continuum.

Procedure. In all, 360 subjects were tested individually. The experimenter spelt out the consonant unit, letter by letter, through a tape-recorder and immediately called the name of the colour or the three-digit number or whatever was the case, whereupon, the subjects started naming the colours or reading the three-digit numbers as they appeared at a constant rate on the memory drum. The subjects engaged themselves in respective intervening activities till they heard the auditory signal "yes" from the tape-recorder at the end of some stipulated retention interval, whereupon they attempted to recall and write down the originally heard three-consonant unit. Response latency was measured by means of a stop-watch from the "yes" signal to the completion of the written response. Two sorts of data were obtained from each subject: (*a*) correct or incorrect reproduction and (*b*) latency of recall.

RESULTS AND DISCUSSION

Recall. The mean recall scores (number of consonant trigrams correctly recalled) for the four principal experimental groups of subjects with interpolated nonsense syllables, three-digit numbers, perceptual discrimination of lengths of lines, and colour naming were, respectively, 5·54, 7·53, 9·18, and 9·95 across the three different time intervals.

All observed differences in these recall means except that between the last two mean recall scores were statistically significant beyond 0·01 level of confidence. The significance of difference between these means was evaluated by *t*-tests. The mean scores of the three retention interval (6, 15 and 24 sec.) groups were respectively, 7·62, 7·15 and 9·38 (explanation of this enhanced score is given below) across the four different kinds of interpolated materials. Mean recall scores under each one of the 12 experimental conditions are given in Table I.

TABLE I

MEAN RECALL SCORES OF THE EXPERIMENTAL GROUPS WITH DIFFERENT INTERPOLATED ACTIVITIES AT DIFFERENT RETENTION INTERVALS

Interpolated activity	*Retention intervals*			*Means for kinds of interpolated activity*
	6 sec.	*15 sec.*	*24 sec.*	
Nonsense syllables.. ..	5·63 (30)	4·53 (30)	6·46 (30)	5·54 (90)
Three-digit numbers ..	7·53 (30)	5·83 (30)	9·23 (30)	7·53 (90)
Perceptual discrimination	8·13 (30)	9·07 (30)	10·33 (30)	9·18 (90)
Colour naming	9·17 (30)	9·17 (30)	11·50 (30)	9·95 (90)
Means for retention intervals	7·62 (120)	7·15 (120)	9·38 (120)	Total N=360

Note.—Figures in brackets indicate the number of measures on which the corresponding mean is based.

Analysis of variance of recall scores revealed significant F-ratios for kinds of interpolated activity ($F = 20\cdot0$, $d.f. = 3, 348$, $p < 0\cdot01$) and retention intervals ($F = 10\cdot1$, $d.f. = 2, 348$, $p < 0\cdot01$).

This shows that differences in the kinds of materials filling the retention interval affect the amount of recall. The original hypothesis that interpolated nonsense syllables would impair recall most, and three-digit numbers next, has been largely borne out by the results. The rather dissimilar interpolated tasks of perceptually discriminating between the lengths of lines and colour naming, impaired recall decisively less.

Some studies underestimate the effect of similarity (Brown, 1958; Posner and Konick 1966) emphasizing the overall difficulty (strenuousness?) of the intervening task. However, reading three letters of the nonsense syllable and reading three digits of the number from the memory drum in the present experiment seem equally strenuous and yet the more similar interpolated material produced more decline in recall.

Latency: The average latencies of correct recalls under interpolated nonsense syllables, three-digit numbers, perceptual discrimination of lengths of lines, and colour naming were, respectively, 4·19, 4·08, 3·81 and 3·55 sec. Mean latencies of incorrect recalls under the same conditions were 5·08, 5·12, 4·25 and 4·33 respectively. It may be noted that, in this connection, only completed incorrect responses were taken into account. It is interesting to note that the latencies of incorrect recalls are uniformly larger than those of correct performances. Apparently, even while an incorrect response was completely executed, it was not without some sense of blocking, conflict, hesitation, etc. (indicators of interference?).

The total number of overt intrusions in recall were 289 with interpolated nonsense syllables and only 16, 32, and 12 with numbers, discriminated lines, and colours respectively, which provides additional evidence for the interference hypothesis of short-term forgetting and is in line with Underwood's (1957) observation that "Amount of interference from other tasks is closely tied to similarity."

A very disturbing factor in this, as in many other experiments on short-term memory, is the possibility of uncontrolled rehearsal during interpolation, to which may be ascribed the paradoxical finding (mentioned earlier) that the longest retention interval gave the largest recall scores.

It was felt that from the point of view of a more effective elimination of rehearsal, a better controlled design of filling the retention interval "with material relevant to the ongoing serial task itself" (Posner, 1963) was employed by Waugh and Norman (1965) in their probe-digit experiment which, incidentally, supported interference as the main source of forgetting. It was, therefore, decided to analyse further the similarity–dissimilarity dimension in its effect on short-term retention using the general procedure of the above-mentioned experiment. In the following experiment, similarity was restricted to acoustic similarity.

EXPERIMENT 2

The hypothesis to be tested more specifically in this experiment was that short-term recall level will be relatively low with acoustically similar words and relatively high in the case of acoustically dissimilar words. For this purpose, recall of minimally rehearsed words was tested and compared after an equal number of other acoustically similar or dissimilar words (and, therefore, also equal number of sec.) had followed their presentation.

Materials. Series of 16 words were used based on Baddeley's (1966*a*) List A comprising 10 acoustically similar words (man, cab, can, cad, cap, mad, max, mat, cat, map) and List B comprising an equal number of acoustically dissimilar words (pit, few, cow, pen, sup, bar, day, hot, rig, bun) matched for Thorndike–Lorge (1944) frequency of occurrence in English language. Sixteen-item word lists were prepared from these two sets with the help of standard tables of random numbers, with the constraint that no word should occur more than twice in the same sequence. No sequence contained words from more than one set. The final word in each series was one that had appeared exactly once before in Position 3, 5, 7, 9, 10, 11, 12, 13 or 14. Upon its reappearance this final "probe-word" was to serve as a cue for the recall of the verbal unit that had followed it previously in the sequence.

The similar and different sequences were recorded on two separate magnetic tapes at a uniform rate of one word per sec. Each of the above-mentioned nine probe-word positions was tested 10 times; each tape thus had 90 test series plus 8 practice series. The final word in each series, the probe-word, was accompanied by a note of 435 ∼ per sec. to help the subject in locating the end of the word series. The position of the first presentation of the probe-word was changed randomly from sequence to sequence but the presentation order for both the tapes was the same.

Procedure. This was a group experiment. The subjects were required to write down the word that came after the probe-word, guessing if they did not remember it exactly. Maximum recall time allowed was 10 sec. As the probe-word occurred only once before in the list, there was only one possible correct answer in each test. There were 12 experimental sessions, six with each tape—the similar and dissimilar tapes occurring in ABBA sequence. The entire first session of similar and dissimilar tapes, each comprising 98 tests, and also the first eight series of every session, were practice trials which, unknown to the subjects, were not scored. The subjects were explicitly and repeatedly told not to rehearse any of the previous words but to think only of the last one heard. By these instructions, it was thought that we were, in effect, measuring the recall of unrehearsed pairs of words, under the two experimental conditions of similar and dissimilar words.

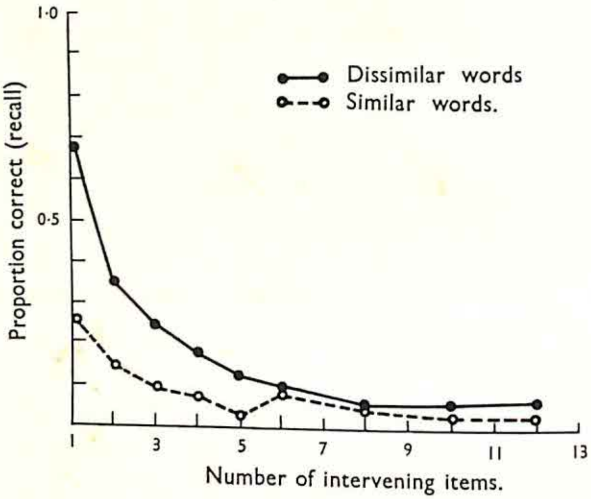
The subjects were six post-graduate students.

RESULTS AND DISCUSSION

Responses were scored as correct or incorrect and group as well as individual serial position curves were plotted under each condition (similar and dissimilar), combining relative frequency of a unit's correct recall along y-axis with its distance from the end of the series on the x-axis (which also gives the number of intervening items). An inspection of the two general functions reveals the difference that similarity–dissimilarity makes to the level of short-term retention, under the conditions of this experiment. Pooled data, corrected for guessing, are given in Figure 1.

Analysis of variance was performed on the mean number of items recalled by each subject at different distances from the end (i.e. with differing numbers of intervening items), under the two experimental conditions. The F-ratio for the similar–dissimilar distinction was found to be significant beyond 0·01 point ($F = 70{\cdot}4$, $d.f. = 1$). With equal

FIGURE 1



Results of Experiment 2. (Each point in Fig. 1 is based on 300 observations.)

retention interval and equal number of intervening items, the observed recall differentials must be ascribed to the similarity–dissimilarity characteristics of the interpolated words.

The F-ratio for position (i.e. distance from the end of the series and also indicating the number of intervening items) was significant at the 0·01 level ($F = 43{\cdot}2$, $d.f. = 8$) and so was that for subjects ($F = 4{\cdot}54$, $d.f. = 5$). The only significant interaction was between position and the similar–dissimilar distinction ($F = 9{\cdot}75$, $d.f. = 8$, $p > 0{\cdot}01$).

The result of this experiment agrees with that of Wickelgren (1965) in showing that increasing the similarity of interpolated material to that in store, causes decrement in the level of retention. As in the experiments of Baddeley (1966*b*) acoustic similarity has been found to depress short-term recall of words.

References

Baddeley, A. D. (1966*a*). The influence of acoustic and semantic similarity on long-term memory for word sequences. *Quart. J. exp. Psychol.*, **18**, 302–9.

Baddeley, A. D. (1966*b*). Short-term memory for word sequences as a function of acoustic, semantic and formal similarity. *Quart. J. exp. Psychol.*, **18**, 362–5.

Brown, J. (1958). Some tests of the decay theory of immediate memory. *Quart. J. exp. Psychol.*, **10**, 12–21.

Peterson, L. R., and Peterson, M. J. (1959). Short-term retention of individual verbal items. *J. exp. Psychol.*, **58**, 193–8.

Posner, M. I., and Konick, A. F. (1966). On the role of interference in short-term retention. *J. exp. Psychol.*, **72**, 221–31.

Posner, M. I. (1963). Immediate memory in sequential tasks. *Psychol. Bull.*, **60**, 333–49.

Thorndike, E. L., and Lorge, I. (1944). *The Teacher's word Book of* 30,000 *Words*. New York: Columbia University Press.

Underwood, B. J. (1957). Interference and forgetting. *Psychol. Rev.*, **64**, 49–60.

Waugh, W. C., and Norman, D. A. (1965). Primary memory. *Psychol. Rev.*, **72**, 89–104.

Wickelgren, W. (1965). Acoustic similarity and retroactive interference in short-term memory. *J. verb. Learn. verb. Behav.*, **4**, 53–61.

Manuscript received 23rd April, 1968.

BOOK REVIEWS

Annual Review of Psychology, Vol. 19, 1968. Edited by P. R. Farnsworth., J. Polefka and M. R. Rosenzweig. Palo Alto: Annual Reviews Inc. Pp. xii + 544. $9·00.

This year's *Review* contains nine articles on the following subjects: Audition, Visual Acuity, Brain Functions, Secondary Motivational Systems, Computer Simulation, Verbal Learning and Memory, The Learning of Psychomotor Skills, Developmental Psychology, Personality Structure, Mass Communication, Psychotherapeutic Processes, Statistical Theory, Consumer Analysis, Psychology of Men at Work, Student Development and Counselling. It offers for the first time a reprint service for individual articles published in the *Review* since 1961, each costing 95 cents.

Reading straight through the *Review* serves as a stimulating and salutary reminder of the breadth of what is taken to be psychology; and the styles of the articles are as varied as their content. One can only be grateful that anyone should be prepared to undertake the often mammoth task of selection from the literature, a task which in most subjects grows each year. Despite the enormity of this task, alluded to somewhat nervously by most of the authors, and despite the obvious need for it, it is never difficult to criticize *Annual Reviews*, and this one is no exception. It seems in parts both over-ambitious and pedestrian, too detailed and naive. This dissatisfaction arises at least partly from ambiguity about the precise functions, possible or intended, of such *Reviews*. One is constantly driven to ask exactly what interests are meant to be catered for; neither the editor nor any of the authors explicitly address themselves to this question, leaving one unsure as to how to judge the articles. As instant history of science, they cannot but fail. The work reviewed is too recent, the involvement of the authors in it too great, for the necessary detachment and theoretical perspective. The only chapter in this volume which seems (to a complete layman) to come anywhere near success by these criteria is that on Computer Simulation by Earl Hunt. It may well be that the requirements of this particular subject are such that one cannot even begin to study it without making explicit its logical foundations, something that unfortunately does not seem to be true of the rest of psychology, as represented in this volume. Perhaps a little research should be done on who uses the *Annual Reviews*, and for what purposes. One assumes that they are intended for those wishing to find out something about subjects that are not in their major or current fields of interest, since they come out too long after the initial publication of papers to keep any specialist up to date. The articles are clearly written to be something more than well selected bibliographies, though it is difficult not to feel, perhaps unfairly, that some of them do not really rise above this level. Perhaps there would be less dissatisfaction if this bibliographic function could, within each article, be explicitly separated from other possible functions, such as providing a readable account, that is not just a catalogue, of "what has happened" in a particular field.

As regards content, it is only possible to comment on those subjects of which one has some first-hand experience, and it would therefore be invidious to name the particular articles, but there are some rather strange omissions. In all cases these are omissions of what seems to the present reviewer to be important European work, including much published in English. In general nearly all the chapters are devoid of references to non-American work, inevitably in some subjects, but making one feel that this volume at least should be more properly called the *Annual Review of* American *Psychology*. There are the usual run of massively uninformative statements, mostly in the more applied subjects, e.g.: "People start smoking due to such factors as conformity, curiosity, enjoyment, and rebellion against authority . . ."; "Research on student development . . . informs us that the college as context for student development is a significant variable as it impinges on students." There is a strange claim to objectivity in the request that the term "consumer solicitude" should replace the allegedly less objective one of "consumer protection." And it might be thought just slightly old-fashioned, in these days of sophisticated genetic analysis, to regard the finding of a strong heritability component in certain psychomotor tasks as grist to a reactionary mill.

Despite these criticisms, it must be reiterated that at times when it is hard enough to keep up with the exploding literature in one's own fields of interest, any form of intelligent selection in other fields is immensely welcome.

JOANNA RYAN.

Selective Attention and Evoked Potentials. By R. Näätänen. (*Ann. Acad, Scientiarum Fennicae*, Ser. B, 151.) Helsinki: Suomalainen Tiedeakatemia. 1967. Pp. 226. £2 10s.

This monograph is divided into two parts. The first part contains a critical assessment of the results of studies of changes in evoked potentials in animal and human subjects during selective attention. No one who has worked in this field escapes criticism, and for the most part the criticisms are sound. Näätänen reviews this work in relation to the hypothesis that evoked responses to unattended stimuli should be smaller than those to attended stimuli. He concludes, rightly I think, that "no evoked potential amplitude correlates for selective perception were provided by these investigations" (p. 43).

The essential point made in the second part of the monograph, which details his own experiments, is that if stimulus A is presented at the time when the subject is required to attend to that stimulus, or to ignore it and attend to stimulus B, the evoked potential to A recorded at the scalp will be *larger* than when the subject is not required to attend to any stimulus. Such changes in the evoked potential therefore are non-specific ("increased reactivity to all kinds of stimuli") rather than specific ("increased reactivity to relevant stimuli and/or diminished reactivity to irrelevant stimuli") (p. 62).

I doubt if these results represent the last word for since this monograph appeared claims have already been made that changes in some of the late components of the evoked potential are "specific". Within another two years it will certainly be necessary to review the field once again.

Certain conclusions which Näätänen draws from his own studies of evoked potentials are not entirely warranted. He writes that these studies " . . . do not support the hypothesis that selective attention and selective perception are based on selective inhibition of irrelevant impulses." It might well be, however, that these experiments are not incompatible with this view. There is often a disassociation between the activity of single cells and of the electroencephalogram, and changes in one may not be reflected in changes in the other. Hence, it is possible for there to be marked changes in the firing rate or responsiveness of single cells, without associated changes in the electroencephalogram.

If selective perception is a real phenomenon, it seems likely that there will be a neural basis for it; and it may be that a study of the slow wave activity of the brain is singularly unsuited to search for and analyse the putative neural mechanism.

G. Horn.

A View of the Brain. By J. S. Griffith. Oxford Science Research Papers 1. London: Clarendon Press; Oxford University Press. 1967. Pp. vi + 79. 15s.

In this book a brave attempt is made to think in very broad terms about the whole nervous system, and to construct mathematical theories; it is well worth reading. But the function of theory should be to predict the outcome of an experiment not yet performed; by this test the book is somewhat disappointing.

The book describes a number of properties of networks already accepted; that facilitation of a synapse leads to instability; that a pool of excitatory neurons can exist in two stable states; that two such pools can interact to produce a limit cycle. Some novel hypotheses are made, and so clearly that they are worthy of consideration, and criticism. One hypothesis is that some as yet undetermined substances pass across a synapse so as to change its properties; but this is asking the question "How does it do it?" before asking the question "What does it do?" A combination of psychology and computer science will one day show how a network can effect pattern recognition. One can then ask a neurophysiologist whether neurons function similarly; and after the latter scientist has told us what neurons do to one another we can go to the biochemist for detailed mechanisms.

The second hypothesis is that there may be very large numbers of different categories of neurons, members of one category being prevented by some chemical means from forming synapses with neurons of more than a few other categories; in this way selected channels of communication could be set up which would not disturb other channels. This is rather like the present state of this planet in which each of us can speak two or three languages of the hundreds which exist. Is not a common language more likely to produce a stable system? Occam must turn in his grave. Excitation and Inhibition are facts, and their interplay can be sufficient to ensure that one interrelated neuron pool does not disturb another.

Lastly, there is the hypothesis of a brain centre and a cortex interacting with another. Such a system would have the property of solving one problem at a time; psychological data support this idea. But where is the experiment? Presumably one should look for some central rhythm out of phase with cortical α rhtyhm. But the author is expressly speaking about non-specific cortical areas, and in these we do not observe α rhythm.

This review is a debate between a micro-theory man and a macro-theory man. There is room and need for both; it is hoped that the above discussion will demonstrate that Professor Griffith's ideas must be examined very carefully.

A. M. UTTLEY.

Human Learning and Memory: Selected Readings. Edited by Norman J. Slamecka. London and New York: Oxford University Press. 1967. Pp. x + 543. 46s.

This expensive paperback contains 50 papers, all by American authors, reprinted in the main from four American journals. (*Journal of Experimental Psychology, Psychological Review, Journal of Verbal Learning and Verbal Behaviour* and *American Journal of Psychology*). The 50 papers are divided into ten sections, covering the interference theory of forgetting, serial learning processes, the serial position effect, continuity versus discontinuity, free recall, short-term memory, transfer, psycholinguistics, concept formation and verbal conditioning. Each section is preceded by some concise remarks by the Editor about the five papers following, together with five additional references "for those who wish to study the problem more extensively."

This book is designed as a basic source of the more frequently cited "key" articles in the area of human verbal learning and memory, and, among other things, as "a practical solution to the problem encountered when a dozen or more students, all seeking the same article, converge simultaneously upon the library." In these respects, the book succeeds. The ten fields chosen by Slamecka are certainly among the most important fields of investigation at the present moment, although perhaps one would have preferred to see serial learning processes and the serial position effect condensed into one section, leaving room for a section devoted to mathematical learning theories. Also, the scope of the sections themselves could have been widened a little; for instance, instead of "the interference theory of forgetting," Slamecka might have considered "possible causes of forgetting." The 50 articles *are* among the more frequently cited, by American psychologists at least (although whether one can equate "more frequently cited" with "key" is arguable), and it is certainly useful to have them collected together in an organized fashion. The question is not whether or not this book is useful, but rather whether its usefulness outweighs its cost. In my opinion, this is not so for individuals, except in the case of a small minority of students who have no means of access to the four Journals mentioned above. However, I think the usefulness of this book does outweigh its cost as far as libraries are concerned.

There are two main ways in which the usefulness of a collection of articles could outweigh its cost to individuals. First of all, if a number of important articles were included which had not been previously published in a widely circulated journal, but had appeared merely in technical reports (etc.), this would make the book worth buying, and also help in the spread of knowledge. For instance, the article by Egan (1958) on the application of signal detection theory to memory, is often cited, but very difficult to get hold of, and the same applies to several technical reports on mathematical learning theories. Secondly, the book might have been especially worth buying if the Editor had written a considerable part of it himself, instead of a very short introduction to each section. For instance, the Editor might have written a concluding chapter raising such questions as why research has moved in these directions, and in which directions research might move in the future. The Editor could also have tried to clarify some of the important issues in learning, for instance the nature of aquisition, retention, retrieval and decision processes, which are often inextricably intermingled in paired-associate learning experiments. Alternatively, the Editor might have considered how results from experiments in the laboratory might be applied (or have been applied) outside the laboratory, and, indeed, might have considered why factors which appear so important outside the laboratory (for instance, motivation) are largely not investigated inside the (human) laboratory, and why some of the phenomena studied so extensively inside the laboratory (for instance, the serial position effect) are largely neglected outside the laboratory.

Since all these articles are fairly easy to get hold of, and the contributions from the Editor are not remarkable, it seems to me that a student of human learning and memory would be better advised to obtain a recent textbook, rather than this collection of readings. However, this book would be very useful as a reference book in the library.

D. P. FARRINGTON.

Cognitive Psychology. By Ulric Neisser. New York: Appleton Century Crofts. 1967. Pp. xi + 351. £3 8s. $7.50.

Cognitive Psychology is a misleading title, suggesting a work on conceptual thinking. Only the last of the eleven chapters is addressed to that topic, and even it includes also memory, imaging, hallucinations etc. The rest of the book is about the perception of visual form and of written and spoken language. Perhaps once in five years in human experimental psychology a book is published which so exactly fills a need in its subject matter, and is so well written that it can be recommended without reservations. If Garner's *Uncertainty and Structure* was the last book of this type, *Cognitive Psychology* is the current one. It is therefore unfortunate that the title does not reflect the content more closely. Neisser's book may even be more influential than Garner's, because it covers a wider field, organizing it in terms of a simple idea, it is aimed at a less advanced or specialized reader, and is hardly concerned with formal problems at all—but an introductory dig at information measurement.

The book has many faults that it does not tax the critic to discover. But this is not the point. Using a book is not an analytic activity but a synthetic one. Neisser's theme, onto which he grafts such superficially disparate phenomena as character recognition and hypnagogic imagery is that of perception as involving an active synthesis of internal representations to be matched with external events. Among other phenomena this idea is consistent with complete perception in the presence of an incomplete stimulus. So it is with reading the book. Because the overall outline of the book so exactly fulfills a need it makes the reader synthesize his own version in which the faults are corrected. A second edition (the first has already reprinted) might correct some of these details as well as incorporating some of the current research important to the theme. The above opinions of the book are not unique, agreeing with the overall reactions of a postgraduate discussion group which devoted four sessions to *Cognitive Psychology*.

This qualified eulogy is intended to encourage the reader to ignore the title and use the book; if he does, he will find chapters on Iconic Storage and Verbal Coding; Pattern Recognition; Focal Attention and Figural Synthesis; Words as Visual Patterns; Visual Memory; Speech Perception; Echoic Memory and Auditory Attention; Active Verbal Memory; Sentences; Higher Mental Processes. "Iconic storage," and "echoic memory" refer in the visual and auditory modalities respectively to the short-term sensory trace that decays in about one or two seconds. The import of the two chapters concerned is that only very basic processes of analysis precede the sensory trace store, and whether the task be recognition or attention extracting the information involves a rapid, synthetic transferral to some more permanent form of storage.

The active synthesis idea is a very old one, although only influential recently as reintroduced by Mackay. Early in the book Neisser makes the necessary distinction between synthesis of an internal representation and actual involvement of motor processes. He shows in the Pattern Recognition chapter that feature analysis cannot explain the flexibility of human pattern recognition, although it is a necessary requirement of it. To account for recognition invariance under transformations such as inversion and to account for segmentation in cursive handwriting or in speech (Hebb's "acquired unity)" some kind of synthesizing selective process has to be invoked. Unfortunately for experimental psychology, as this reviewer has pointed out elsewhere, it is difficult to distinguish between such active mechanisms and passive ones in recognition on the basis of input-output relations alone, and one has to resort to logical arguments and the pseudoscientific principles of neuro-economy; but the knowledge that things *must* be so is never quite as satisfying as the knowledge that they are. This problem remains with the theme throughout the book, although it is not seriously discussed. Neisser uses the "active synthesis" idea as a line on which to peg out the cognitive washing and this is quite valid as long as it is realized that interpretation of the experimental support for "active synthesis" models are in fact based upon an intuitive assessment of what such a model involves rather than on rigorous predictions that it makes. This Titchenerian tendency might be offset by inclusion in the second edition of the recent British work on attempts to make "active" theories generate rigorous predictions.

Neisser appreciates that "synthesis" theories are intimately related to perceptual selectivity and hence attention. But in the chapter on focal attention and figural synthesis, Neisser says that to direct attention to a figure is "to attempt a more extensive synthesis of it" (possibly some circularity in "more extensive"?); now, if synthesis is to be invoked thus to solve the segmentation problem, and the detail-under-transformation problem there must be an accompanying selection of the sub-set of stimulus aspects corresponding to the sub-set of synthesized aspects. Thus synthesis *requires* selectivity and is not identical to it. But in the chapter on auditory attention, Neisser explains away attention as being synonymous with active synthesis rather than any filtering prior to the recognition stage! No account is taken of the fact that most selective attention experiments have explicitly *required* an active synthesis of subjects in the crude sense of the shadowing response. These are the difficulties of working with nice ideas like "synthesis" rather than explicit models. Neisser should have protected the unwary reader by pointing out that the difficulties do exist, and by saying explicitly, for example, that more than one level of selective or synthetic process was possible.

In order to write a book that is useful for varied specialists wishing to keep in touch with an active field (there are about 500 references) Neisser must inevitably have left some stones half-turned. One review paper is cited as "lucid" which this reviewer often cites as an example of a paper that is undesirably loose. Neisser is often under-critical, and each specialist will be able to find unobserved niceties in his own area. But non-specialisation has its good sides as well; Neisser has not worked in what is currently termed psycholinguistics; he points out that the insistence in structural analysis in that field has much in common with Gestalt Psychology. Neisser clearly understands Gestalt Psychology (and Psychoanalysis); this reviewer never did understand it fully, and the parallels drawn have been of some help. In addition the excellent chapter on sentences is much more useful than it probably would have been had Neisser himself been steeped in the current obscurantist tradition in psycholinguistics.

With these virtues of usability and readability does it matter that some methodological and conceptual problems are raised and not worked out? That introspective evidence is used and linked to the "synthesis" model in an informal fashion, or that "demand characteristics" of experiments are used to justify skeptimism only when it suits the train of the argument? Undergraduates will have to be forewarned, but in a world of the Unexceptionable the Good will always be welcome.

MARK P. HAGGARD.

Experiments in Animal Behaviour. By Marguerite D. Hainsworth. London: Macmillan. 1967. Pp. x + 206. 35s.

It appears that animal behaviour is being studied increasingly in Biology courses in schools and Colleges of Education, and this book is addressed mainly to people concerned with such work. It provides detailed instructions for carrying out experiments on a wide range of animals, from protozoa to rodents, with the emphasis on the study of reponse to light, temperature, and chemical and mechanical stimulation. The experiments have been ingeniously planned to exploit the limited technical resources likely to be available to students, and the book is clearly the product of considerable practical teaching experience. It aspires to be more than a laboratory manual, and it is full of suggestions intended to stimulate independent work. But the best of pedagogic intentions are liable to be degraded in the classroom, and one would like to be sure that the study of animal behaviour at this level would not turn into the performance of a set of standardized laboratory exercises. Possibly because each chapter deals with a variety of experiments with a particular group of animals rather than with experiments on a particular behavioural capacity in different species, the book is rather disjointed, and it sometimes reads like a set of hints for amusing the children on wet afternoons during the school holidays.

Would it help University students of Psychology to have pursued at school the kind of course to which this book relates? It would certainly give them a good deal of practical experience in experimenting with animals, but it might be questioned whether this is what is most needed. One of the special problems of animal behaviour as a scientific subject matter is the difficulty we have in understanding the need for explanations that go beyond the level of everyday conversation, and then in constructing such explanations. This is a problem that this book hardly acknowledges.

M. H. SHELDON.

PUBLICATIONS RECEIVED

BOOKS

Studies in Invertebrate Behaviour. By S. M. Evans. London: Heinemann Educational Books. 1968. Pp. vii + 110. 16s.

Social Facilitation and Imitative Behaviour. Edited by E. C. Simmel, R. A. Hoppe and G. A. Milton. Boston: Allyn and Bacon. 1968. Pp. x + 253.

Anxiety and Neurosis. By C. Rycroft. London. The Penguin Press. 1968. Pp. xiii + 159. 30s.

Psychiatrie im Grundriss. By Hans Jorg Weitbrecht. Berlin, Heidelberg and New York: Springer-Verlag. 1968. Pp. xvi + 490. DM.46.

The Perception of People and Events. By P. B. Warr and C. Knapper. London, New York and Sydney: Wiley. Pp. xiii + 445. 75s.

The Scientific Study of Abnormal Behavior. By James Inglis with a foreword by Joseph Zubin. Chicago: Aldine Publishing Co. 1966. Pp. xiv + 296. $8.95.

Courtship: A Zoological Study. By Margaret Bastock. London: Heinemann Educational Books. 1967. Pp. viii + 220. 25s.

The Senses Considered as Perceptual Systems. By James J. Gibson. London: George Allen & Unwin. 1968. Pp. xv + 335. 40s.

Eye Movements and Vision. By A. L. Yarbus. Translation Editor: Lorrin A. Riggs. New York: Plenum Publishing Corp. 1967. Pp. xiii + 222. $17.50.

The Discrimination Process and Development. By Brian J. Fellows. Oxford: Pergamon. 1968. Pp. xii + 218. 75s.

The Experimental Psychology of Sensory Behavior. By John F. Corso. New York: Holt, Rinehart & Winston. 1967. Pp. xii + 628. $12.95.

PAPERBACKS

Conditioning and Instrumental Learning. By Edward L. Walker. Belmont, California: Brooks/Cole Publishing Company. Basic Concepts in Psychology Series. 1967. Pp. xii + 161. $2.35.

Discrimination Learning. By Donald A. Riley. Boston, Mass.: Allyn & Bacon. 1968. Pp. x + 171.

The Pelican History of Psychology. By R. Thomson. London: Penguin Books. Pelican Original. 1968. Pp. 464. 7s. 6d.

Person als Prozess. Festschrift zum 65 Geburstag von Professor Dr. Phil Robert Heiss. Edited by K. J. Goffmann and K. H. Wewetzer. Bern and Stuttgart: Verlag Hans Huber. 1968. Pp. 242. DM.19.80.

Creativity and Personal Freedom. By Frank Barron. London, Princetown, Toronto, New Jersey, Melbourne: Van Nostrand. 1968. Pp. viii + 322. 37s. 6d.

Computer Produced Physiological Tables. By G. R. Kelman and J. F. Nunn. London: Butterworths. 1968. Pp. 50.

Perspectives in Clinical Psychology (*University Series in Psychology*). By A. Z. Guiora and M. A. Brandwin. London, Princetown, Toronto, New Jersey and Melbourne: Van Nostrand. 1968. Pp. xiv + 333. 39s. 6d.

Die Seelishchen Krankheiten des Menschen. By P. C. Kuiper. Bern and Stuttgart: Hans Huber. 1968. Pp. 278. DM.19.80.

Readings in Measurement and Evaluation: Education and Psychology. By N. E. Gronlund. New York and London: Macmillan. 1968. Pp. xv + 463. 45s.

The Organization of Human Information Processing. By F. Klix. Berlin: Akademie-Verlag. 1967. Pp. 195.

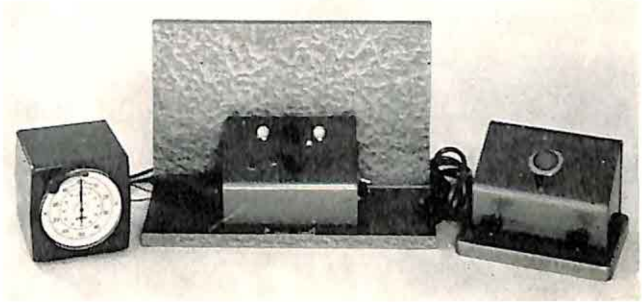
INDEX TO AUTHORS

Agnew, N. McK., with Pyke, S., and Pezer, V. 110
Allen, L. R., with Garton, R. F. 385
Amsel, A., with Surridge, C. T., and Mock, K. R. 321
Baddeley, A. D. 249
Bertelson, P., with Tisseyre, F. 297
Bindra, D., with Corballis, M. C., and Lieberman, W. 51
Bradshaw, J. L. 116
Brindley, G. S., with Wolff, J. G., Delacour, J., and Carpenter, R. H. S. 1
Carpenter, R. H. S., with Wolff, J. G., Delacour, J., and Brindley, G. S. 1
Clark, H. H., with Clark, E. V. 129
Clark, E. V., with Clark, H. H. 129
Coltheart, M., with Von Sturmer, G., and Wong, T. 380
Corballis, M. C., with Lieberman, W., and Bindra, D. 51
Corcoran, D. W. J., with Weening, D. L. 83
Corcoran, D. W. J., with Dorfman, D. D., and Weening, D. L. 336
Cosgrove, K., with Hershberger, W. A., and Trantina, P. R. 395
Day, R. H., with Wade, N. J. 290
Delacour, J., with Wolff, J. G., Carpenter, R. H. S., and Brindley, G. S. 1
Dodwell, P. C., with Kuechler, H. A. 305
Dorfman, D. D., with Corcoran, D. W. J., and Weening, D. L. 336
Dornbush, R. L. 225
Fischer, G., with Viney, W., Knight, J., and Johnson, N. 301
Fisher, G. H. 212
Garner, W. R., with Gottwald, R. L. 97
Garton, R. F., with Allen, L. R. 385
Geffen, G., with Treisman, A. 139
Glencross, D. J., with Provins, K. A. 282
Gottwald, R. L., with Garner, W. R. 97
Green, F. A., with Templeton, W. B. 200
Grindley, G. C., with Townsend, V. 11
Hardwick, J., with Shaffer, L. H. 360
Hardy, G. R., with Legge, D. 20
Harter, M. R., with White, C. T. 157
Haslam, D. R. 208
Hershberger, W. A., with Trantina, P. R., and Cosgrove, K. 395
Houlihan, K., with Sekuler, R. W. 204
Hughes, R. N. 189
Jamieson, J. L., with Wong, R. 197
Jeffrey, H. J. 294
Johnson, N., with Fischer, G., Viney, W., and Knight, J. 301
Johnson-Laird, P. N. 69
Jones, S. 74
Kahneman, D., with Onuska, L., and Wolman, R. E. 309
Karlin, L., with Kestenbaum, R. 167
Kennedy, R. A., with Wilkes, L. 390
Kestenbaum, R., with Karlin, L. 167
Knight, J., with Fischer, G., Viney, W., and Johnson, N. 301

Kothurkar, V. K. - 405
Kuechler, H. A., with Dodwell, P. C. - 305
Legge, D., with Hardy, G. R. - 20
Lester, D., with Still, A. W. - 86
Lieberman, W., with Corballis, M. C., and Bindra, D. - 51
Likely, D., with Schnitzer, S. B. - 193
Millodot, M. - 329
Mock, K. R., with Surridge, C. T., and Amsel, A. - 321
Murdock, B. B. - 79
Onuska, L., with Kahneman, D., and Wolman, R. E. - 309
Pezer, V., with Agnew, N. McK., and Pyke, S. - 110
Pizzurro, S. A., with Van Laer, J. - 151
Pritchatt, D. - 351
Provins, K. A., with Glencross, D. J. - 282
Pyke, S., with Agnew, N. McK., and Pezer, V. - 110
Rabbitt, P. M. A. - 179, 232 and 241
Sanders, A. F., with Schroots, J. J. F. - 370 and 373
Savin, H. B. - 123
Schnitzer, S. B., with Likely, D. - 193
Schroots, J. J. F., with Sanders, A. F. - 370 and 373
Sekuler, R. W., with Houlihan, K. - 204
Shaffer, L. H. - 38
Shaffer, L. H., with Hardwick, J. - 360
Sheldon, M. H. - 400
Still, A. W., with Lester, D. - 86
Surridge, C. T., with Mock, K. R., and Amsel, A. - 321
Templeton, W. B., with Green, F. A. - 200
Tisseyre, F., with Bertelson, P. - 297
Townsend, V., with Grindley, G. C. - 11
Trantina, P. R., with Hershberger, W. A., and Cosgrove, K. - 395
Treisman, A., with Geffen, G. - 139
van Laer, J., with Pizzurro, S. A. - 151
Venables, P. H., with Warwick-Evans, L. A. - 30
Viney, W., with Fischer, G., Knight, J., and Johnson, N. - 301
von Sturmer, G., with Wong, T., and Coltheart, M. - 380
von Wright, J. M. - 62
Wade, N. J., with Day, R. H. - 290
Warwick-Evans, L. A., with Venables, P. H. - 30
Wason, P. C. - 273
Weening, D. L., with Corcoran, D. W. J. - 83
Weening, D. L., with Corcoran, D. W. J., and Dorfman, D. D. - 336
White, C. T., with Harter, M. R. - 157
Wilkes, L., with Kennedy, R. A. - 390
Wolman, R. E., with Kahneman, D., and Onuska, L. - 309
Wong, R., with Jamieson, J. L. - 197
Wong, T., with von Sturmer, G., and Coltheart, M. - 380
Wolff, J. G., with Delacour, J., Carpenter, R. H. S., and Brindley, G. S. - 1
Wright, P. - 265

Model RT-5 Visual Choice Reaction Time Apparatus

Four Stimuli
Separable Consoles
Monitor Lights
"Get Ready" Signal



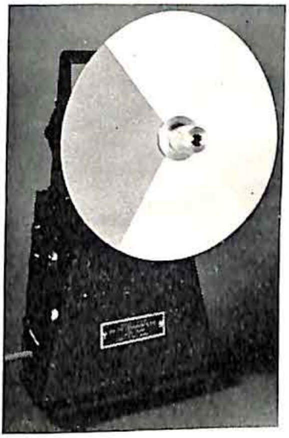
Measures Simple,
Discriminative,
and Choice
Reaction Times

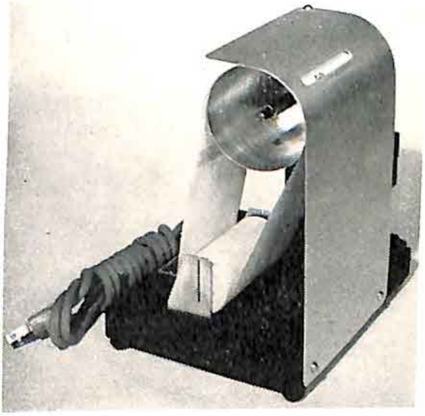
Harvard Tachistoscope

With solid state,
direct reading timer
Available in two and three field models

Colour Mixer

Heavy base
Quality Bodine motor
Easy to carry
Has series type speed control
Will handle discs up to 17″ in dia
A rugged unit suitable for student and research use





Memory Drum

Our model M1-A is of SIMPLE, DURABLE design
Has 3 exposure speeds, can be used with long or short word lists, is an ideal drum for use in class or laboratory

Electronic Voice Key

All solid state Model C-370

Complete catalogue available on request

Ralph Gerbrands Company

8 Beck Road, Arlington, Mass. 02174 U.S.A. (617) 648-6415

Psychologische Forschung

Basic Research in Psychology

Edited by
O.-J. Grüsser, Berlin
R. Held,
Cambridge, Mass.
Th. Herrmann, Marburg
H. Hörmann, Berlin
H.W. Leibowitz,
University Park, Pa.
P. Leyhausen,
Wuppertal-Elberfeld
D. Ploog, München
K.-H. Wewetzer,
Giessen

"Psychologische Forschung" invites contributions in general experimental psychology. Inasmuch as psychology is broadly concerned with problems of behavior and conscious processes, contributions from related behavior-oriented disciplines, such as zoology, ethology, physiology, psychiatry, sociology, and neurology are encouraged — provided they contribute to our understanding of basic psychological problems.

All Contributions in English beginning with Volume 31:

Ingle, D., G. Schneider, C. Trevarthen and R. Held: Locating and Identifying: Two Modes of Visual Processing. A Symposium

Ingle, D.: Two Visual Mechanisms Underlying the Behavior of Fish

Schneider, G. E.: Contrasting Visuomotor Functions of Tectum and Cortex in the Golden Hamster

Battacchi, M. W., M. Chiarenza, G. Giovanelli and P. Tampieri: Character Impulsivity in Terms of Speech and Motor Behavior

Michon, J. A.: A Model of Some Temporal Relations in Human Behavior

Trevarthen, C. B.: Two Mechanisms of Vision in Primates

Held, R.: Dissociation of Visual Functions by Deprivation and Rearrangement

Schöne, H., and H.-G. Mortag: Variation of the Subjective Vertical on the Parallel Swing at Different Body Positions

In the interest of speedy publication this journal is issued at frequent intervals, according to the material received. As a rule 4 numbers constitute one volume. The price is DM 120,—; US $30.00 per volume. Members of the "American Psychological Association" are granted a 15% price reduction.

■ **Prospectus on request!**

Springer-Verlag
Berlin - Heidelberg - New York